

**A Thesis Submitted for the Degree of PhD at the University of Warwick**

**Permanent WRAP URL:**

<http://wrap.warwick.ac.uk/104204>

**Copyright and reuse:**

This thesis is made available online and is protected by original copyright.

Please scroll down to view the document itself.

Please refer to the repository record for this item for information to help you to cite it.

Our policy information is available from the repository home page.

For more information, please contact the WRAP Team at: [wrap@warwick.ac.uk](mailto:wrap@warwick.ac.uk)

# Essays On Bayesian Persuasion

Davit Khantadze

A thesis presented for the degree of Doctor of  
Philosophy  
Department of Economics  
University of Warwick  
United Kingdom  
September, 2017

# Contents

<b>1</b>	<b>Literature review</b>	<b>7</b>
1.1	Introduction . . . . .	7
1.2	Different approaches to bayesian persuasion . . . . .	8
1.2.1	Choosing optimal posterior distributions - Con- cavification approach . . . . .	8
1.2.2	Information design approach to bayesian persua- sion . . . . .	13
1.3	Competition in Persuasion . . . . .	14
1.4	Informed receiver(s) . . . . .	18
1.5	Dynamic bayesian persuasion . . . . .	20
1.5.1	Ely: Beeps . . . . .	20
1.5.2	Hörner and Skrzypacz: Selling Information . . . .	25
1.6	Further contributions to bayesian persuasion . . . . .	28
1.7	Multidimensional persuasion . . . . .	29
<b>2</b>	<b>Two-dimensional bayesian persuasion</b>	<b>30</b>
2.1	Introduction . . . . .	30
2.2	Related work . . . . .	33
2.3	The model . . . . .	34
2.3.1	Payoffs . . . . .	34
2.3.2	Signal structures . . . . .	34
2.3.3	Sequential signal structure and order of moves . .	35
2.3.4	Simultaneous signal structure and order of moves	36
2.3.5	Optimal signals . . . . .	36
2.4	General Observations . . . . .	36

2.5	Optimal simultaneous and sequential signal structures, when $n=1$ . . . . .	41
2.5.1	Optimal simultaneous signal . . . . .	42
2.5.2	Sequential signal structure . . . . .	44
2.6	Characterising conditions for a sequential signal struc- ture to achieve an upper bound when $n=2$ . . . . .	51
Appendix 2.A	Optimal signals when $n=2$ and only one de- cision is to be made . . . . .	52
<b>3</b>	<b>Identifying the reasons for coordination failure in a labora- tory experiment</b>	<b>57</b>
3.1	Introduction . . . . .	57
3.1.1	Related works . . . . .	60
3.2	Example . . . . .	64
3.3	The model . . . . .	65
3.4	Experimental design . . . . .	67
3.4.1	Predictions . . . . .	69
3.4.2	Hypotheses . . . . .	71
3.5	Results . . . . .	73
3.6	The role of beliefs in the fight against female genital mu- tilation . . . . .	79
3.7	Conclusion . . . . .	82
Appendix 3.A	Belief hierarchies . . . . .	84
Appendix 3.B	Equilibrium selection and models of higher- order beliefs . . . . .	84
Appendix 3.C	Order effects . . . . .	86
Appendix 3.D	Data . . . . .	89
Appendix 3.E	Online Appendix . . . . .	90
3.E.1	Instructions . . . . .	90
3.E.2	Quiz . . . . .	94
<b>4</b>	<b>Bibliography</b>	<b>98</b>

# Acknowledgments

I want to thank my advisors Ilan Kremer and Motty Perry for their guidance. I am especially grateful to Ilan Kremer for encouragement and insightful discussions.

I also would like to thank Kirill Pogorelskiy who read my paper.

I am grateful to various faculty members at the economics department at Warwick for stimulating conversations.

I would like to thank Philipp Külpmann for being an excellent co-author.

Financial support from the economics department at Warwick and ESRC studentship made my stay at Warwick possible. I appreciate this.

I want to dedicate this work to my family.

# Declaration

I declare that all the following work was carried out by me. Except the following:

Chapter 3 I co-authored with Philipp Külpmann.

# Abstract

Chapter 1 reviews the literature about the bayesian persuasion. It first describes two main approaches to bayesian persuasion: concavification approach and information design approach. Next I consider some extensions to the basic model of bayesian persuasion, like competition between different senders, privately informed receiver and dynamic bayesian persuasion. Some other contributions reviewed include costly bayesian persuasion and bayesian persuasion when receiver's optimal action is only a function of an expected state.

Chapter 2 deals with two-dimensional bayesian persuasion. In this chapter I investigate a model when the receiver has to make two decisions. I am interested in optimal signal structures for the sender. I describe the upper bound of sender's payoff in terms of his payoff when only marginal distributions of two dimensions are known. Completely characterise optimal simultaneous and sequential signal structures, when each dimension has binary states. This approach extends concavification approach to bigger state space, than explored in previous contributions to bayesian persuasion. Finally I characterise optimal sequential signal structure when there are three states for each dimension.

In chapter 3 I investigate together with my co-author the effect of absence of common knowledge on the outcomes of coordination games in a laboratory experiment. In our experiment, around 76% of the subjects have chosen the payoff-dominant equilibrium strategy despite the absence of common knowledge. However, 9% of the players had first-order beliefs that lead to coordination failure and another 9% exhibited coordination failure due to higher-order beliefs.

# Chapter 1

## Literature review

### 1.1 Introduction

Economists aim to solve optimisation problems; the latter can take different forms. In the current paper we want to discuss a class of papers that broadly deal with bayesian persuasion, i.e. how one party, the sender, can provide information optimally to the other party, the receiver. Optimality means maximising sender's payoff. The intention of the literature on bayesian persuasion is to understand situations, where one party has to make a decision based on the information provided by the other party. Examples might be following: seller can provide some information to the buyer about the product quality; the prosecutor undertakes investigation about the crime and provides this information to the judge; politician provides information about the quality of the project to the public. The question becomes interesting if there is some conflict of interest between the sender and the receiver. The seller might want to sell the product independent of its quality, whereas the buyer wants to buy the product only if the quality is good. The prosecutor might want to send the defendant into the prison independent if the latter is guilty or not, whereas the judge wants to declare guilty only if the defendant is indeed guilty.

This review intends to discuss some main contributions in the field and can not claim to give the complete review of the bayesian persuasion literature. First, in section 1.2 we discuss two different approaches



to bayesian persuasion; first approach, made popular by Kamenica and Gentzkow [2011], uses the concavification argument, as developed by Aumann and Maschler [1995]; second approach uses the concept of bayesian equilibrium and extends it to the case where the mediator has informational advantage relative to the players of the game. This approach was pioneered in a series of papers by Bergemann and Morris [2016a,b]. Next, in section 1.3 we discuss the case when multiple senders try to persuade a single receiver. Section 1.4 reviews literature about informed receiver, i.e. when receiver has some additional information. In section 1.5 we review two papers that deal with dynamic setting. Last section briefly reviews some other approaches to bayesian persuasion.

## **1.2 Different approaches to bayesian persuasion**

In this section we discuss how economists think about acquiring information by sender, when he does not have commitment problems. First we explore the concavification approach.

### **1.2.1 Choosing optimal posterior distributions - Concavification approach**

Let's consider following problem: receiver has to make a decision -  $a$ . Sender has preferences over receiver's actions. Receiver's optimal action is a function of his beliefs -  $\hat{a}(\mu)$ . For example judge's decision is a function of his belief about defendant being guilty or innocent. Say sender can choose, given some common prior  $\mu_0$ , receiver's posteriors. What is the optimal way for him to do this? This question is dealt by Kamenica and Gentzkow [2011] and we intend to reproduce here some of their main arguments.

Receiver's utility is a function of his action and of the state of the world -  $u(a, \omega)$ , where  $a \in A$  is receiver's action and  $\omega \in \Omega$  state of

the world. The knowledge about state of the world is given by common prior distribution  $\mu_0 \in \Delta(\Omega)$ . Sender can influence receiver's beliefs by choosing a signal. Signal is a mapping  $\pi : \Omega \rightarrow \Delta(S)$ . It means that sender can choose conditional distributions given state,  $\pi(\cdot|\omega)$ . Probability distribution is over some signal realisation space  $S$  and  $s_i$  is some element in  $S$ . While it might not be immediate how to address the problem of choosing optimal information structure, concavification argument says that there exists a simple solution to this problem if one is able to draw the graph of sender's value function as a function of receiver's beliefs. Before further exploring this question let's consider following example, that was formulated by Bergemann and Morris [2016a] and is equivalent to the example by Kamenica and Gentzkow [2011].

A depositor has to decide to stay with the bank ( $s$ ) or not ( $r$ ). State of the world is good ( $G$ ) or bad ( $B$ ) and bank is solvent if  $\omega = G$  and insolvent if  $\omega = B$ . Depositor and regulator have common prior and assign equal probability to both states. Depositor's utility function is the following:

$$\begin{aligned} u(s, G) &= y & u(s, B) &= -1 \\ u(r, G) &= 0 & u(r, B) &= 0 \end{aligned} \tag{1.1}$$

We assume that  $y \in (0, 1)$  in general. For this example  $y = 0.5$ .

Regulator prefers depositor to stay with the bank, independent of the state of the world. Regulator's utility if depositor chooses ( $s$ ) is 1 and 0 if depositor chooses ( $r$ ). While it might not be immediate what is the optimal signal for the regulator, analysing regulator's problem geometrically, makes solution straightforward.

Before doing this, first we want to summarise the concavification approach, as formulated by Kamenica and Gentzkow [2011].

Let's state one observation, that will turn out helpful in our search for optimal signal structure.

From the law of total probability it follows that for arbitrary signal, following must be true:

$$\mu(\omega) = \mu(\omega|s_1)p(s_1) + \dots + \mu(\omega|s_n)p(s_n) \quad (1.2)$$

Equation 1.2 says that for any signal expected posterior must equal to prior. It turns out that this is the only constraint that a distribution of posterior beliefs has to satisfy.

Following observation is important for the concavification approach: instead of thinking about a particular signal, one can think about the distribution of posterior beliefs, that satisfy equation 1.2. We reproduce Kamenica and Gentzkow [2011]'s result here.

**Proposition 1.** *Following statements are equivalent:*

- i There exists a signal that gives sender some expected payoff  $v^*$ ;*
- ii There exists a distribution of posteriors, whose expectation equals to prior and sender's expected payoff equals  $v^*$ .*

Let's denote sender's value function by  $v(\mu)$  and its convex hull by  $co(v)$ . Now we want to introduce the concept of concavification of the function. Given prior  $\mu$ , sender's maximum feasible payoff can be described in the following way:

$$V(\mu) \equiv \sup \{x_2 | (\mu, x_2) \in co(v)\} \quad (1.3)$$

$V(\mu)$  is the smallest concave function everywhere weakly greater than  $v(\mu)$  and Aumann and Maschler [1995] refer to  $V(\mu)$  as the concavification of  $v(\mu)$ .

Now we can summarise concavification approach to finding sender's optimal signal structure:

- 1 : Draw sender's value function as a function of posterior beliefs -  $v(\mu)$
- 2 : Draw the concave closure of  $v - V(\mu)$
- 3 : Find distribution of posteriors  $\mu_s - (\tau)$ , s.t.  $E_\tau v = V(\mu)$
- 4 : Given  $(\tau)$ , calculate signal structure that induces  $(\tau)$ .

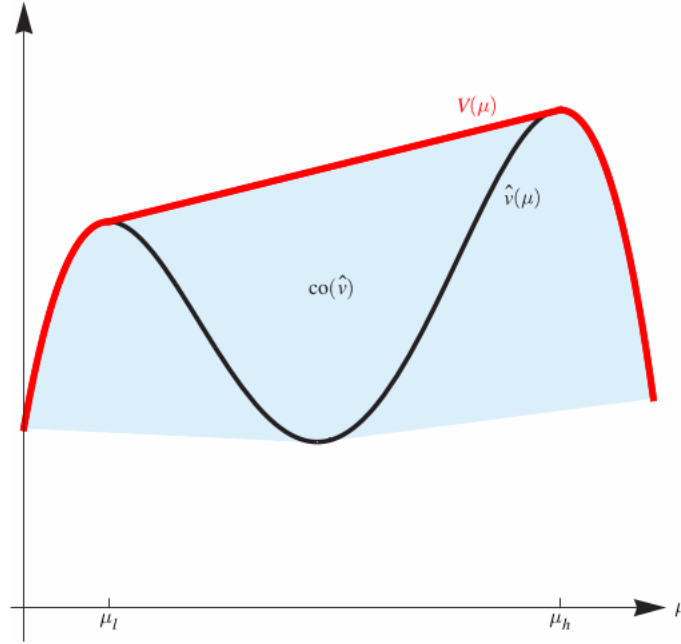


Figure 1.1: Concavification of a function - an example; copied from Kamenica and Gentzkow [2011]

Kamenica and Gentzkow [2011] show that step 4 is always possible.

We will use this approach to find regulator's optimal signal.

First we want to construct regulator's value function as a function of depositor's beliefs. If  $\mu(\omega = G) < \frac{2}{3}$  then  $\hat{a}(\mu) = r$  and if  $\mu(\omega = G) \geq \frac{2}{3}$ , then  $\hat{a}(\mu) = s$ . Thus regulator's value function, without signals, is the following:

$$0 \quad \text{if} \quad \mu(\omega = G) < \frac{2}{3} \quad (1.4)$$

$$1 \quad \text{if} \quad \mu(\omega = G) \geq \frac{2}{3} \quad (1.5)$$

Now, by using equation 1.2 we are able to describe all possible expected payoffs regulator can achieve by choosing some signal structure. Regulator's feasible expected payoffs can be described in the following way:  $x = (0.5, x_2) \in co(v)$ . This is the vertical line in figure

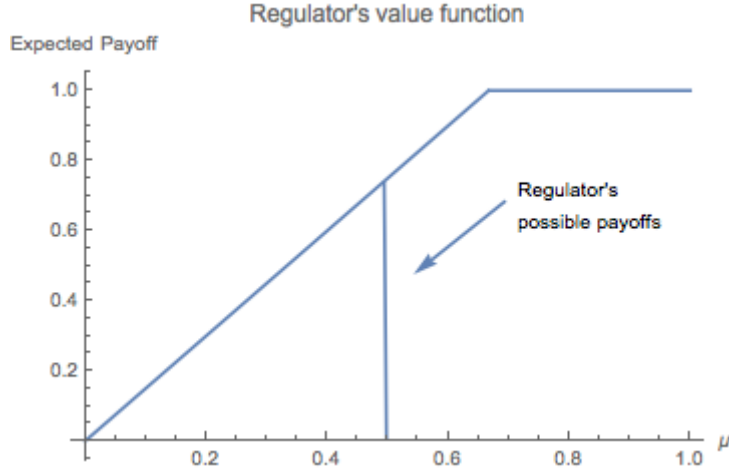


Figure 1.2: Concavification of regulator's value function

1.2.

Concavification of regulator's value function is the following:

$$V(\mu) = \min\left\{\frac{3}{2}\mu, 1\right\} \quad (1.6)$$

Optimisation implies that regulator would choose highest available expected payoff. One can see that this payoff is given by the following formula:  $\lambda \frac{2}{3} + (1 - \lambda)0 = \frac{1}{2}$ , where  $\lambda$  is the probability that posterior equals  $\frac{2}{3}$ . So, regulator can choose a signal structure that induces following distribution of posterior beliefs:

$$\mu = \frac{2}{3} \quad \text{with probability} \quad \frac{3}{4} \quad (1.7)$$

$$\mu = 0 \quad \text{with probability} \quad \frac{1}{4} \quad (1.8)$$

What is the signal structure that gives this distribution of posteriors? Note that bayes rule, together with equation 1.7 implies that  $\pi(s_s|\omega = G) = 1$ . Therefore following signal structure induces desired distribution of posteriors:

$$\begin{aligned}
\pi(s_s|\omega = G) &= 1 & \pi(s_s|\omega = B) &= \frac{1}{2} \\
p(s_r|\omega = G) &= 0 & p(s_r|\omega = B) &= \frac{1}{2}
\end{aligned} \tag{1.9}$$

Now we have accomplished the following: we described a relatively simple procedure how to find a joint distribution of signal realisations and states that is most favourable for the regulator.

	$\omega = G$	$\omega = B$
$s_r$	0.50	0.25
$s_s$	0.00	0.25

The goal of the exercise was to show that when one can graph the value function of the sender as a function of beliefs, then one can find optimal signal structure by first finding optimal distribution of posterior beliefs.

Now we describe information design approach to bayesian persuasion, as suggested by Bergemann and Morris [2016a,b].

## 1.2.2 Information design approach to bayesian persuasion

Bergemann and Morris [2016a,b] develop an information design approach, that uses the concept of mediator, which gives recommendations possibly to multiple receivers. They show that if there is a single receiver, then their approach to information design reduces to "bayesian persuasion". As we will see, the information design approach makes it relatively easy to model the private information of the receiver, i.e. when the receiver also observes additional, possibly public, informative signal about the state of the world. Before discussing the case of the receiver with additional information, we will discuss the above example by using the concepts as suggested by Bergemann and Morris [2016a,b].

We want to think about the problem as discussed in subsection (1.2.1) in the following way: regulator makes action recommendations

given state; therefore we say that mediator has an information advantage. Let's denote by  $\rho_\theta$  regulator's recommendation for the depositor to run, when the state is  $\theta \in \{B, G\}$ . Now, regulator's recommendation has to satisfy obedience constraint, i.e. it has to be optimal for the depositor to follow the recommendation. Therefore, the depositor will have an incentive to stay, if

$$(1/2)(1 - \rho_G)(1/2) - (1/2)(1 - \rho_B) \geq 0, \quad (1.10)$$

and an incentive to run if

$$0 \geq (1/2)\rho_G(1/2) + (1/2)\rho_B(-1). \quad (1.11)$$

We note that obedience constraint for staying 1.10 implies obedience constraint for run 1.11.

Therefore, we can write 1.10 in the following form:

$$\rho_B \geq 1 - \frac{1}{2} + \frac{1}{2}\rho_G \quad (1.12)$$

Say regulator's goal is to minimise the probability of depositor running, then we see from inequality 1.12 that  $\rho_G = 0$  and  $\rho_B = \frac{1}{2}$ . Note that this recommendation rule will make the depositor stay with probability  $\frac{3}{4}$ . Note also that recommendation rule and signal structure as given in the expression 1.9 coincide.

After having described general approaches to bayesian persuasion, now we will discuss contributions that extend bayesian persuasion as described above in several dimensions. First we will discuss the case of several senders, i.e. when there is a competition between senders.

### 1.3 Competition in Persuasion

Our discussion of competition in persuasion is based on the following contributions by Gentzkow and Kamenica [2016a, 2017a].

The main question of Gentzkow and Kamenica [2016a] is the following: what is the property of information structures that guarantees

that competition among senders will increase the amount of information revealed. To begin with, Gentzkow and Kamenica [2016a] construct an example where competition among firms reduces the amount of information revealed. It turns out that there is a simple characterisation of information structures that guarantees that competition does not decrease the amount of information. Competition can not decrease the amount of information revealed if and only if following is true: every player can induce any feasible distribution of posteriors that is more informative than posteriors induced by all other senders. One distribution of beliefs is more informative than other, if it is a mean preserving spread of the other. Gentzkow and Kamenica [2016a] call this condition Blackwell-connected. So, if information structure is not Blackwell-connected one can construct an example when competition can decrease the amount of information revealed.

One also would like to know what is the equilibrium outcome when there is competition among senders. To be able to describe equilibrium outcomes one has to assume that all senders have access to the same set of signals. Call a distribution of beliefs unimprovable for sender  $i$ , if for any other feasible distribution of beliefs sender  $i$  would not be better off. Then, Gentzkow and Kamenica [2016a] are able to prove following result:

**Proposition 2.** *Say information environment is Blackwell-connected and each sender has access to the same set of signals. Then a feasible outcome is an equilibrium outcome if and only if it is unimprovable for each sender.*

One can see that if an outcome is not unimprovable, then Blackwell-connected means that there exists a sender  $i$  that can increase his payoff by inducing a more informative distribution of beliefs, that will increase his payoff. From this characterisation of the set of equilibrium outcomes it follows that one can easily find this set: one has to find the intersection of unimprovable outcomes. Unimprovable outcome can be described in the following way: the support of the distribution of posteriors  $\mu$  is such that  $v(\mu) = V(\mu)$ , where  $v(\mu)$  is sender's expected utility without signal and  $V(\mu)$  is concavification of sender's value function. Kamenica and Gentzkow [2011] show that for one sender,



optimality implies this condition. But with many senders, Gentzkow and Kamenica [2016a, 2017a] show that this condition should be satisfied for all senders.

We will illustrate this by the following example: Graph (a) depicts value function of sender 1 and graph (b) of sender 2. Solid line intervals indicate unimprovable beliefs for each sender. The set of equilibrium outcomes is the intersection of these sets and in this example is the following:  $[\tau_1, \tau_2] \cup \{1\}$ .

Horizontal axis in the graph depicts the share of facts revealed, so 1 means that all facts are revealed; therefore moving to the right of the horizontal axis means revealing more information.

Graph (d) depicts the collusive outcome, i.e. revealed facts that maximise senders 1 and 2 overall payoffs. As one can see all equilibrium outcomes reveal more facts than collusive outcome. This is in accordance with the claim that if information environment is Blackwell-connected then collusive outcome can not reveal more information.

One can also see that fully informative outcome is an equilibrium. This follows from proposition 2. So, if all senders choose from the same set of signals and information environment is Blackwell-connected, then most informative feasible outcome is always an equilibrium.

We could apply these ideas to our example about regulator and depositor. Say there also exists another, competing bank that prefers depositor to run than to stay. So it gets payoff 1 if the depositor runs and 0 if it stays. Let's call it competitor. It can also produce signals for the depositor. Assume that depositor stays if his posterior of bank being good is at least  $\frac{2}{3}$ . Then concavification of competitor's value function is given in figure 1.4.

So, when regulator chooses his optimal signal, then competitor could reveal information about the bank and increase its payoff. The only equilibrium outcome of this game is fully informative signals. Note that when fully informative signals are produced, then for any posterior belief  $\mu$ , following is true for both regulator and competitor:  $v_i(\mu) = V_i(\mu)$ ,  $i \in \{\text{regulator}, \text{competitor}\}$ , i.e. beliefs are unimprovable for both senders.

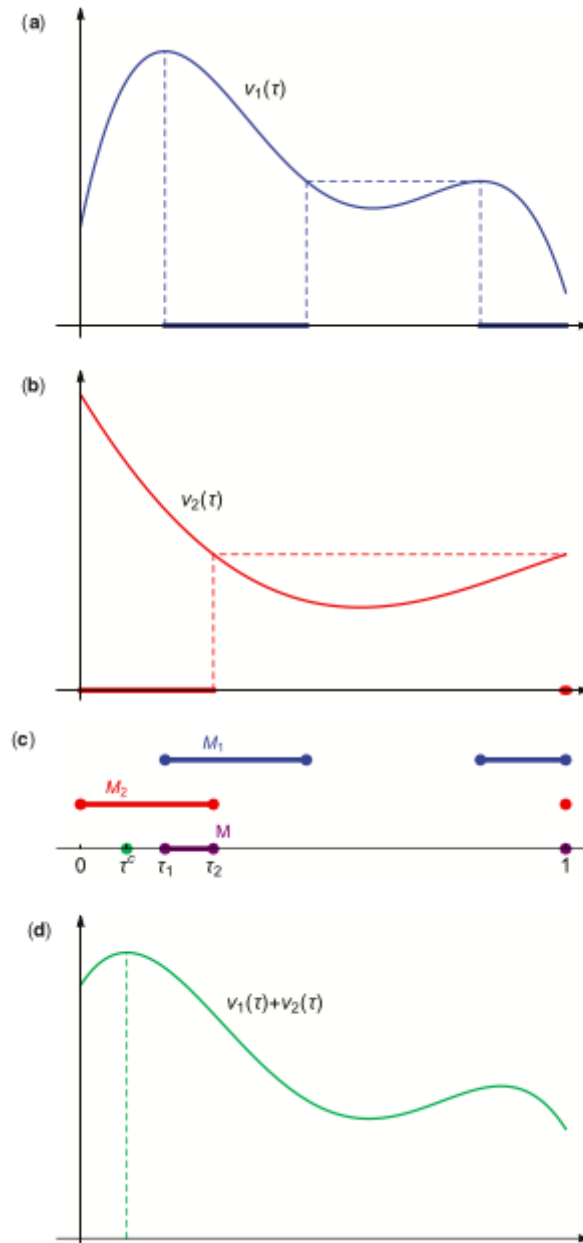


Figure 1.3: Characterising equilibrium outcome. (a) Sender 1's preferences. (b) Sender 2's preferences. (c) Equilibrium construction. (d) Collusive preferences - copied from Gentzkow and Kamenica [2016a]

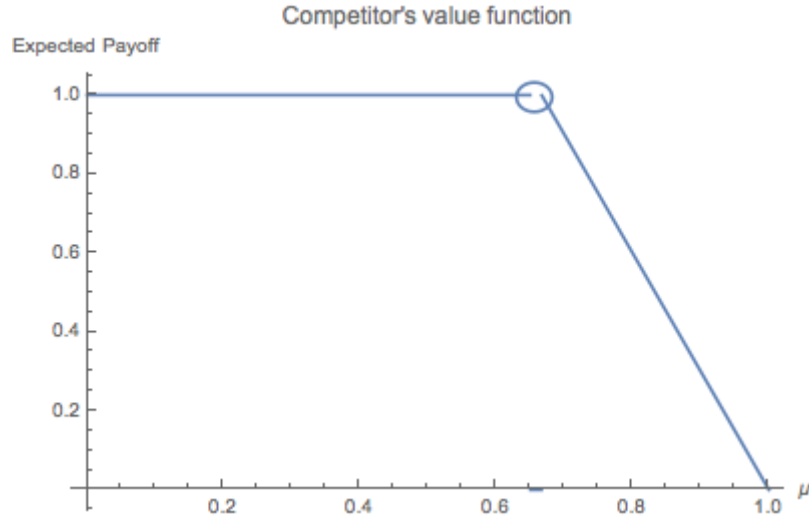


Figure 1.4

This example shows that sender can be worse off from competition, because receiver gets more information. Here receiver's additional information is an outcome of strategic interaction between competing senders. Next we want to review literature that explores following question: receiver gets informative public signal about the state of the world.

## 1.4 Informed receiver(s)

Bergemann and Morris [2016a] divide information design in two steps: first, describe the set of feasible outcomes in general and second describe what outcomes will be chosen from an interested party and what is the information structure that leads to these outcomes.

These approach was already used in section (1.2.2) while discussing information design approach to bayesian persuasion. Now we will illustrate it for the case when the depositor observes informative public signal about the state of the world.

Public signal is drawn from the following joint distribution of the signal space and the decision relevant state space:

	$\omega = G$	$\omega = B$
$g$	$\frac{q}{2}$	$\frac{1-q}{2}$
$b$	$\frac{1-q}{2}$	$\frac{q}{2}$

$$q > \frac{1}{2}.$$

The set of feasible outcomes will also be determined by the following: the regulator observes depositor's public signal or not. First let's consider the case when the public signal, i.e. depositor's additional information, is observed by the regulator. Therefore, regulator's recommendation will be a function of the state and the signal observed.

Now for a recommendation to stay to satisfy obedience constraint, i.e. for depositor to have an incentive to follow the recommendation, it should take into account depositor's conditional belief about the state of the world that he forms after observing the public signal. So, when the depositor observes good signal, then he will follow the recommendation to stay, if

$$q(1 - \rho_{Gg})(1/2) - (1 - q)(1 - \rho_{Bg}) \geq 0 \quad (1.13)$$

Similarly, depositor will follow the recommendation to run after he observed the good signal, if:

$$0 \geq q\rho_{Gg}(1/2) + (1 - q)\rho_{Bg}(-1). \quad (1.14)$$

Similar conditions have to be satisfied if the public signal is bad. In expectation, for a recommendation to be followed, following condition should be satisfied, that we give in a form of proposition:

**Proposition 3.** *The probabilities  $\rho_B, \rho_G$  form an equilibrium outcome for some information structure if*

$$\rho_B \geq \max\{q(1 + \frac{1}{2}), 1\} - \frac{1}{2} + \frac{1}{2}\rho_G. \quad (1.15)$$

This describes the set of all possible recommendation rules that satisfy obedience constraint. As the precision of the public signal ( $q$ ) increases,  $\rho_G$  becomes 0 and  $\rho_B$  becomes 1, so the only feasible recommendation is the fully informative one. Since depositor becomes fully

informed about the state of the world from observing the public signal, there is not much the regulator can do by providing additional information.

Bergemann and Morris [2016a] show that the set of feasible outcomes shrinks when the depositor gets additional information and this information is also observed by the regulator. This set shrinks even more if depositor's additional information is not observed by the regulator and he has to elicit it from the depositor.

Bergemann and Morris [2016b] extend the result that receiver's additional information reduces the set of feasible outcomes to many players case. They identify the condition of "more informative" in the sense of Blackwell in the case of many players, for which the result extends to the many player setting.

Kolotilin et al. [forthcoming] consider following question: receiver has private information about his preferences; sender can commit to information disclosure about the state. Sender's and receiver's utilities are linear function of the state and receiver's type. Sender, before choosing a signal structure could ask the receiver his type and the signal could be a function of the receiver's report. Or sender could choose a signal structure independent of receiver's report. Kolotilin et al. [forthcoming] show that in their setting there is no loss of generality in considering only signal structures. Therefore sender can ignore more complex persuasion mechanisms and produce signal without asking the receiver to report his type.

Now we will discuss some papers that deal with bayesian persuasion in the dynamic context.

## **1.5 Dynamic bayesian persuasion**

### **1.5.1 Ely: Beeps**

Ely [2017] considers dynamic setting of information provision. State of the world evolves and agent's belief changes over time even if the principal does not provide any information. Information provided by the

principal changes agent's current belief and also shapes the path of it's further evolution. Ely [2017] extends the approach to information provision as formulated by Aumann and Maschler [1995] and Kamenica and Gentzkow [2011] to the dynamic setting. In particular he extends the insight that it is without loss of generality to analyse distribution of posterior beliefs, s.t. expectation equals prior, to the dynamic setting and one can use geometric approach to find optimum distribution of posteriors and optimal signal that induces it.

Ely [2017] considers following motivating example. An agent has to work, but is distracted by checking emails. The IT-department (principal) considers how to filter information so that the agent checks the email as late as possible. Principal could use the email notification software that beeps when the email arrives. If the software is switched on, then the worker stops working when he hears the beep. E-mail arrival is modelled as a poisson process. The e-mail arrival rate is  $\lambda$ . So, if the software is switched on, then expected arrival time of the e-mail is  $\frac{1}{\lambda}$ .

Principal could decide to switch-off the software. Assume that the agent has a threshold belief  $p^*$ , s.t. when his belief is at least as big as  $p^*$ , then he checks the e-mail. When the beep is turned off, then agent's belief that an email has arrived is  $1 - e^{-\lambda t}$  and therefore agent's working time is the following:

$$t^* = -\frac{\log(1 - p^*)}{\lambda} \quad (1.16)$$

By comparing equation (1.16) with  $\frac{1}{\lambda}$  one can see that for high  $p^*$  agent will work longer if beep is turned off. If  $p^*$  is low enough then agent works longer when beep is turned on.

These two regimes are two extremes of information disclosure: when beep is turned on, then principal chooses full information disclosure, since the signal perfectly informs about the state of the world, whereas if beep is turned off, then no information is disclosed. Can the principal do better? Does there exist a way to make the agent work longer than any of these two mechanisms? Ely [2017] suggests following

mechanism, that turns out to be also optimal: beep with a delay. If e-mail arrives at date  $t$ , the agent hears beep at date  $t + t^*$ . Expected time of work, as induced by this mechanism, is  $t^* + \frac{1}{\lambda}$ .

Note that the suggested mechanism accomplishes the following: when principal's and agent's interests are aligned, i.e.  $t \leq t^*$ , principal does nothing. When  $t > t^*$ , i.e. agent's belief, without information, is such that he would check the e-mail, then principal filters information such that agent's belief is either  $p^*$  or 1. Note the resemblance to the depositor case, where the depositor learns the true state only if the state is bad. Similarly in the current example the agent learns that email has arrived only if  $t > t^*$ , i.e. he would check the e-mail if there was no additional information provided by the principal.

How does the suggested mechanism accomplish this? Because the agent hears the beep with delay  $t^*$ , it means that he does not hear beep for the period of  $t^*$ . Say  $t > t^*$  and the agent has not heard the beep. Then the agent knows that e-mail has not arrived  $t^*$  periods before and his belief is  $p^*$ .

Before giving the proof that this is principal's optimal mechanism, we briefly describe how to extend the insight from the static bayesian persuasion that it is without loss of generality to consider distribution of posterior beliefs, whose expectation equals prior. In the dynamic context Ely [2017] shows that it is without loss of generality to directly choose a stochastic process for the agent's beliefs, given that this process satisfies two properties: expected posterior equals prior and agent's belief evolves according to the state transition probability.

Ely [2017] first guesses principal's optimal strategy, which is delayed beep. The value function for this strategy, for agent's current belief  $\mu_t$  being smaller than the critical value, i.e.  $\mu \leq p^*$ , is given by the following equation:

$$V(\mu) = \frac{1}{r} \left[ (1 - e^{-r\tau(\mu)}) + e^{-r\tau(\mu)} \left( \frac{r}{r + \lambda} \right) \right] \quad (1.17)$$

For  $\mu \in (p^*, 1]$  guess is made that value function is linear. There-

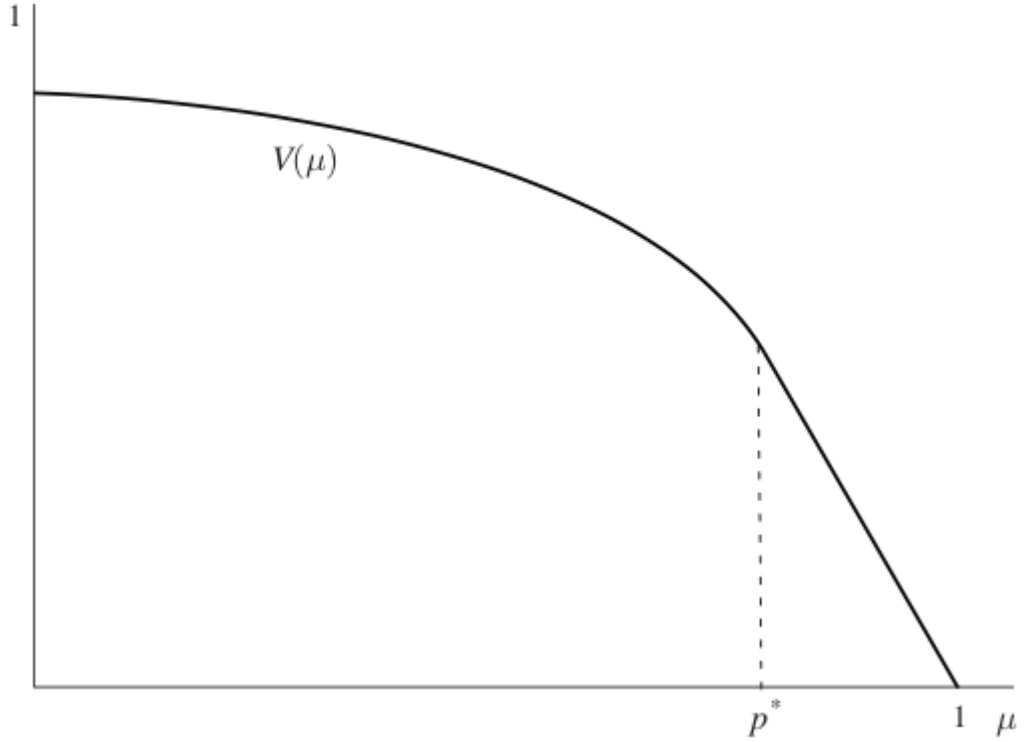


Figure 1.5: Value function for the  $t^*$ –delayed beep - copied from Ely [2017]

fore, the guessed value function is given by figure 1.5

Proving optimality consists in showing that the suggested strategy is unimprovable. By taking into account that feasible policies induce distribution of posterior beliefs whose expectation equals prior belief, one has to consider deviations that satisfy this condition. But then, given this constraint, one can describe the optimal value as a concavification of the value function. Ely [2017] verifies that principal's optimal payoff from one-shot deviation coincides with the suggested value function arising from the strategy of delayed beep.

Optimal payoff can be formulated in the following form:

$$rV = cav \left[ u + V' \cdot \frac{dv}{dv} \right] \quad (1.18)$$



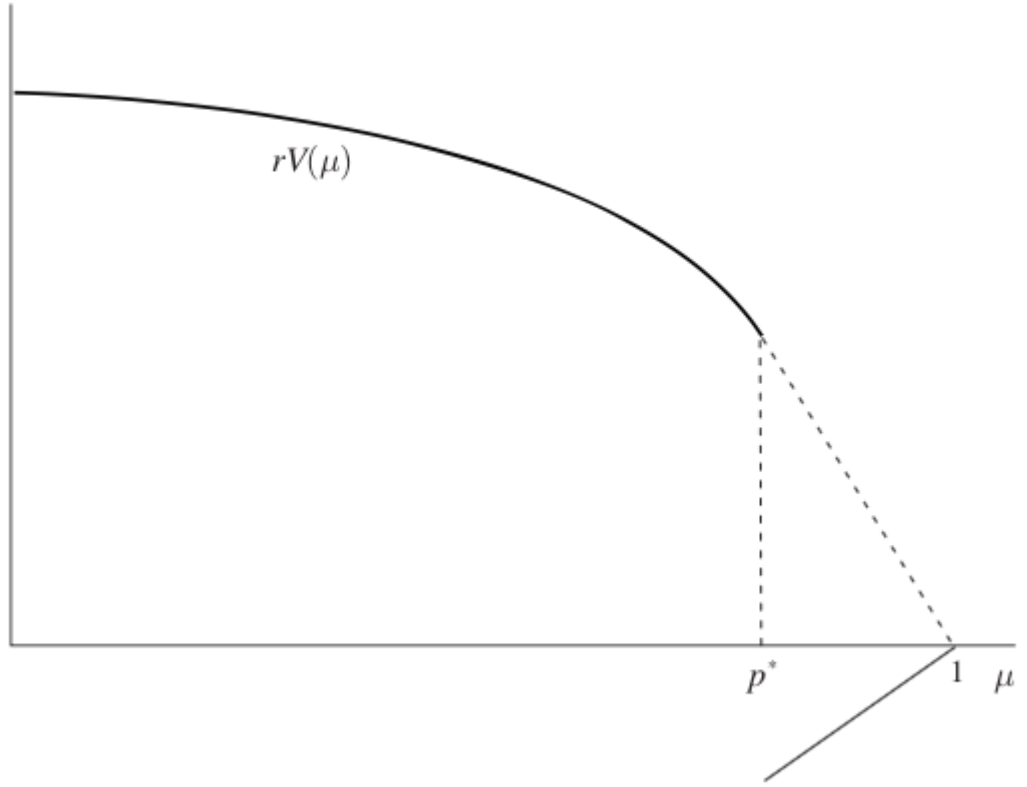


Figure 1.6: Concavification - copied from Ely [2017]

where,  $\text{cav}$  denotes concavification of the expression in the bracket. Graph of the expression in the bracket of equation (1.18) is the following:

Concavification argument also explains why delayed beep is an optimal strategy. The value function without signal is concave for  $\mu \leq p^*$ , therefore it is optimal for the principal to not reveal any information and this is accomplished when beep is delayed by  $t^*$ . When agent's belief becomes bigger than  $p^*$ , then concavification is strictly higher than the value function without signal. In this case the induced posterior beliefs are either  $p^*$  or 1 and this is accomplished by the suggested strategy.

Next we consider a paper by Hörner and Skrzypacz [2016] that deals with the following question: how should the agent split information optimally to overcome a hold-up problem.

### 1.5.2 Hörner and Skrzypacz: Selling Information

Hörner and Skrzypacz [2016] analyse following situation: firm (it) considers to hire an agent (she) to implement a project. Agent can be of good (1) or bad (0) type. It pays off for the firm to implement the project only if the agent is of good type. Agent knows her type. Before the firm makes hiring decision, agent and firm can communicate in several stages; during each communication stage transfers can be made and agent can reveal information about her type by choosing a test. Hörner and Skrzypacz [2016] are interested in the following question: how can a competent agent persuade the firm to hire her and still be rewarded for the competence? Motivation for this question is to understand incentives of the agent for acquiring the competence. Formally, authors are interested in finding best possible equilibrium outcome for the competent agent. They show that agent increases her payoff by revealing information gradually. Before reproducing the argument for gradualism, we briefly describe the set-up, that should help the exposition.

The firm decides to hire the agent and implement the project, only if his belief that agent is competent is above some threshold, denoted by  $p^*$ . Firm's belief determines his outside option, that we denote by  $w(p)$ . If  $p > p^*$ , then firm's outside option is positive, otherwise it is 0. We assume that  $p_0 < p^*$ , i.e. firm's prior belief is smaller than the threshold. Firm's payoff from hiring a competent agent is 1; therefore from firm's perspective, when he beliefs that agent is competent with probability  $p$ , expected overall surplus is  $p$ .

Prior of hiring decision agent and firm communicate for  $K$  stages.

The agent can choose the difficulty of the test in each communication stage. Denote test's difficulty by  $m \in [0, 1]$ .  $m$  is the probability with which the test is passed by the incompetent agent. Competent agent passes the test with probability 1.

The choice of test-technology leads to a distribution of posteriors that equals to prior. So, if  $m$  is such that in case the test is passed, posterior becomes  $p'$ , then one knows also distribution of posteriors, because expectation of this distribution should be prior.

Firm's expected gain, when the agent chooses test-technology, is given by the following equation:

$$E_F[w(p')] - w(p) \quad (1.19)$$

where  $E_F[\cdot]$  is firm's expectation operator.

Hörner and Skrzypacz [2016] consider an equilibrium, where in each communication stage the firm pays it's entire expected gain from the additional information to the agent.

Now we are ready to show why gradual information provision benefits the agent. Say there is 1 round of communication.

By the argument as formulated in proposition (1), Hörner and Skrzypacz [2016] look at the distribution of posterior beliefs. Say agent chooses some test that, if it is passed, leads to a posterior belief  $p_1 \geq p_0$ . From the law of total probability then follows that posterior becomes  $p_1$  with probability  $\frac{p_0}{p_1}$ . By using this information, equation (1.19) becomes

$$\frac{p_0}{p_1}w(p_1) - w(p_0) \quad (1.20)$$

Because we are considering the case when  $p_0 < p^*$ , therefore  $w(p_0) = 0$ , i.e. firm does not implement the project. One can show that  $\frac{w(p_1)}{p_1}$  is increasing in  $p_1$ , therefore one can see that with one stage communication maximum the competent agent can achieve is to choose a perfectly informative test, so that  $p_1 = 1$ . The firm would be willing to pay  $p_0$  for such a test, i.e. with one stage communication competent agent can get ex-ante expected surplus.

But the agent can increase her payoff by offering two tests. First test is chosen such that  $p_1 = p^*$ . Remember that  $w(p^*) = 0$ . Therefore the firm's willingness to pay for this test is 0. Given new posterior  $p^*$ , then by repeating the argument of one stage communication, second test is fully informative and agent's payoff is  $p^*$ .

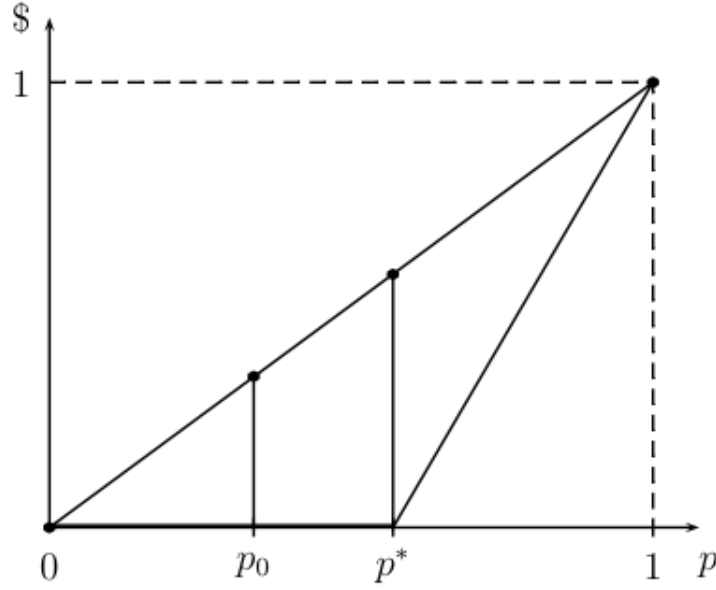


Figure 1.7: Revealing information in two steps - copied from Hörner and Skrzypacz [2016]

Figure 1.7 visualises this argument. Diagonal denotes expected surplus.  $w(p)$ 's curve is below diagonal line and equals 0 for  $p < p^*$ .

Can the agent do better than this by offering more tests? Following example shows that the answer to this question is positive.

Consider now the case when the competent agent offers three tests that lead to the following posteriors:  $p^*$ ,  $p'$  and 1.

Figure 1.8 illustrates agent's payoff from offering three tests. Agent's overall payoff one gets by summing two red vertical lines. One can see that by offering three tests agent can get higher payoff than by offering only two tests. Agent's payoff when only two tests are offered is given by summing the left red line segment and the dashed line segment above it. But as one sees the right red line segment is bigger than the dashed line segment.

Hörner and Skrzypacz [2016] show that when  $p_0 < p^*$ , then agent's payoff is maximised by first giving information for free and making posterior equal to  $p^*$  and then dividing the interval  $[p^*, 1]$  into smaller and smaller intervals.

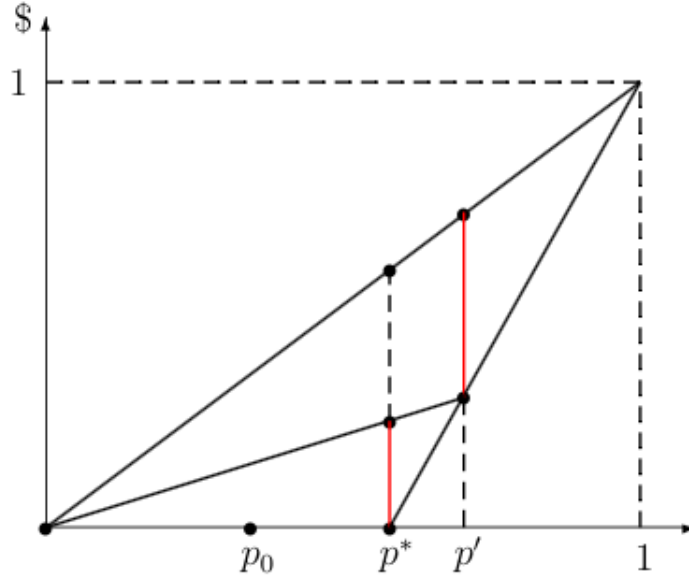


Figure 1.8: Revealing information in three steps - copied from Hörner and Skrzypacz [2016]

Now we briefly survey some other contributions to bayesian persuasion.

## 1.6 Further contributions to bayesian persuasion

Concavification approach requires that sender's value function can be expressed as a function of receiver's beliefs. Gentzkow and Kamenica [2014] consider the case when producing signal is costly. They characterise the class of cost functions which is compatible with concavification approach. Thus, Gentzkow and Kamenica [2014] extend concavification approach to the case when signals are costly.

Gentzkow and Kamenica [2017b] analyse bayesian persuasion (with possibly) many senders, where all senders first privately observe signal realisation. Authors show that endogenous information will be always disclosed and therefore disclosure requirements do not affect equilibrium outcome.

Gentzkow and Kamenica [2016b] considers the setting where receiver's optimal action is a function of expected state and sender's payoff depends only on receiver's action. Authors characterise the set of distribution of posterior means that can be achieved by some signal.

## **1.7 Multidimensional persuasion**

Tamura [2014] considers multidimensional persuasion. It applies semidefinite programming approach to characterise an upper bound for sender's payoff and derives optimal signal when state is normally distributed.

## Chapter 2

# Two-dimensional bayesian persuasion

### 2.1 Introduction

A software company (sender) wants to sell two products (A and B) to the customer (receiver). Sender has to decide how to provide information to maximise the likelihood of receiver buying two products. Sender can choose precision of the information about the products. What is the optimal way to accomplish this when the information about one product also contains indirectly information about the other product?

For illustration, we consider the following example: each product can be good (1) or bad (0). Receiver gets utility 1 if he buys the good product or does not buy the bad product and 0 otherwise. Receiver's utility is additively separable in two products. So, if receiver buys both products and each is good, then his utility is 2. Sender's utility is also additively separable in two products and his utility is 1 if receiver buys a product and 0 otherwise. Sender and receiver share common prior belief about the quality of two products, that is given in figure 2.1.

Sender can design tests that reveal information about the quality of the product. For example sender could allow receiver to test the product and thus learn its quality. Tests are costless and can be of any

	$B = 1$	$B = 0$
$A = 1$	0.30	0.15
$A = 0$	0.05	0.50

Figure 2.1: Motivating Example

precision, i.e. test could inform perfectly about the product(s) or not inform at all.

If  $A$  and  $B$  were independent, then solution to the problem is well known. One has two separate persuasion problems and optimal signal for each product can easily be found by using concavification argument as developed by Aumann and Maschler [1995] and further explored by Kamenica and Gentzkow [2011]. It is well known that concavification approach makes it easy to find optimal signal when the state space is relatively small and it is not straightforward how to use this approach when state space becomes large. This is so, because it relies on geometric argument and requires visualisation to find an optimal signal. This is pointed out for example by Gentzkow and Kamenica [2016b], when they mention the following: "... the value function and it's concavification can be visualised easily only when there are two or three states of the world." Therefore, for a state space bigger than this, it is not immediate how to use concavification of the value function in order to find optimal signal. Thus note that for a simplest possible, non-trivial two-dimensional bayesian persuasion problem it is not immediate how to use concavification approach to find an optimal signal of the sender.

If the sender decides to split the persuasion problem into two parts and first inform about one product and then about the other, he also has to take into consideration the fact that the signal about one product also informs about the other product.

In the current example if the sender decides to inform first only about product  $B$  and then, given new posterior, inform about product  $A$ , one can show that sender's payoff would be smaller than if only marginal distributions of  $A$  and  $B$  were known. But if he chooses to inform first about  $A$  and then about  $B$  then there exists a signal which



gives sender the same payoff as what he would get if only marginal distributions were known. But can the sender do better than this? And does there always exist a signal that informs separately about two products that achieves for the sender the same payoff as when only marginal distributions are known?

We show for the preference specification of our model that the additional information in the form of joint distribution acts as an additional constraint for the sender and for arbitrary number of states for each dimension, sender can never achieve a higher payoff than what he would get if only marginal distributions were known.

From this follows for the current example that there exists a simple procedure for the sender to maximise his payoff: first inform about product  $A$  and then inform about product  $B$ .

Why does the signal that first informs about  $B$  and then about  $A$  fail to achieve the upper bound? Because it reveals too much information about product  $A$ . We describe for arbitrary number of states and for the given preferences of the current model, the necessary and sufficient condition for signals that inform about products separately to achieve the upper bound. This condition states that the support of the distribution of posteriors of one product, as induced by the signal of the other product, should belong to the interval, on which the concavification of sender's value function for this product is linear.

Next we show that when there are two states for each dimension and if  $A$  and  $B$  are positively correlated then there always exists a signal that informs about products separately and achieves the upper bound of the payoff for the sender.

We also show by giving an example that there exist joint distributions, for which there does not exist a signal that informs about two products separately and achieves the upper bound.

The problem of informing about products separately is that the signal about one product might reveal too much information about the other product. Sender can solve this problem by constructing a more complicated signal that informs about both products simultaneously.

When there are two states for each dimension, we construct the

simultaneous signal that informs about both products simultaneously and achieves the upper bound for arbitrary joint distribution.

Next we analyse the case when there are three states for each dimension. Finding a signal that informs about both dimensions simultaneously now means to find a joint distribution of signal space and decision relevant space that has up to 81 states. Currently we do not have solution for this problem. We describe a procedure of how to split this problem into two parts and inform about products separately. In particular we clarify when does this procedure achieve the upper bound for the sender.

The paper is organised in the following way: in the next section we briefly review the literature about bayesian persuasion. Then we describe our model. In section 2.4 we derive some general results. Section 2.5 analyses the case when there are two states for each dimension. Then we analyse optimal sequential signals when there are 3 states for each dimension.

## 2.2 Related work

Kamenica and Gentzkow [2011] analyse optimal information provision when the single receiver has to make a single decision. They give a characterisation of optimal signals in a general framework, by using the concavification argument as formulated by Aumann and Maschler [1995]. This approach can be summarised in the following way: choosing a signal that maximises sender's payoff is equivalent to choosing optimal distribution of posteriors that equals prior in expectation. This distribution of posteriors can be found by drawing value function of the sender as a function of beliefs. After one has found optimal distribution of posteriors, one can also find signal that gives this posterior distribution, by using Bayes rule.

Somewhat related to the current project is the work of Bergemann and Morris [2016a,b], who develop a general approach of information design with multiple receivers, where the latter can also get private signals. Receiver's private, exogenous signal constrains senders at-

tainable outcomes. In the current setting, when the sender chooses to produce information sequentially, receiver's additional information is controlled by the sender and is endogenous in this sense, but still can constrain sender's achievable outcomes.

## 2.3 The model

### 2.3.1 Payoffs

The receiver faces two decision problem,  $A$  and  $B$ , which sometimes we will refer to as dimensions 1 and 2. For each dimension the receiver wants to match the states of the world, which are non negative integers. Receiver's utility function is  $-(a_A - \omega_A)^2 - (a_B - \omega_B)^2$ , where  $a_i, \omega_i \in \{0, 1, \dots, n\}, i \in \{A, B\}$  and some given  $n$ .  $a_i$  denotes receiver's action and  $\omega_i$  denotes the state of the world for dimension  $i$ . We will denote joint state by  $\omega_{A,B} = (\omega_A, \omega_B)$ , where  $\omega_{A,B} \in \{0, 1, \dots, n\}^2$ . Sender's utility function is  $a_A + a_B$ . So, the receiver wants to match the state for each dimension, whereas the sender prefers the receiver choosing as high a number as possible for each dimension.

### 2.3.2 Signal structures

Receiver and sender have common prior joint distribution of  $A$  and  $B$ . Sender can choose a signal structure, which means, choosing a family of conditional distributions. The choice of the signal structure is common knowledge.

The sender could decide to inform about  $A$  and  $B$  separately, which means first producing signal say for  $A$  and then for  $B$ . We call this a sequential signal structure. Or the sender could decide to produce signal for both dimensions simultaneously, i.e. choose a simultaneous signal structure.

If the sender decides to inform about  $A$  and  $B$  separately, then we assume that he can choose the order of persuasion, i.e. sender can choose the dimension about which to produce the first signal.

### 2.3.3 Sequential signal structure and order of moves

A sequential signal structure informs immediately only about one dimension. For simplicity here we assume that the first signal is produced for  $A$  and the second signal for  $B$ . Signal can be viewed as a probability distribution over recommendations, which in equilibrium will be followed by the receiver.

Here we describe a sequential signal structure when the first signal is produced for  $A$ . Sender chooses a family of conditional distribution functions  $\pi(\cdot|\omega_{Ai})$ , over recommendations  $s_{Ai} \in S$ , where  $i \in \{0, 1, \dots, n\}$ .  $S$  denotes the space of signal realisations. The choice of signal for  $A$  is common knowledge. Receiver observes signal realisation  $s_{Ai}$  and updates his beliefs about  $A$  by bayes rule.  $s_{Ai}$  can be interpreted as a recommendation for the receiver to choose  $i$  for  $A$ . In equilibrium this recommendation will be followed.

After observing signal for  $A$ , receiver updates his beliefs about  $A$  and takes an action for  $A$  that is optimal given his beliefs about  $A$ .

Because the common prior is joint distribution of  $A$  and  $B$ , if  $A$  and  $B$  are not independent and if the signal is not uninformative, then the signal for  $A$  will also contain some information about  $B$ . Thus, after updating beliefs about  $A$ , receiver's beliefs about  $B$  might also change.

Given these new beliefs about  $B$  the sender chooses a signal for  $B$ , which is again a family of conditional distributions over recommendations for  $B$ . Receiver observes the choice of the signal for  $B$  and the signal realisation. Then updates his beliefs about  $B$  and takes an optimal action for  $B$ .

When discussing sequential signal structures, we will refer to the signal, say of  $A$ , as optimal, if for the sender this would be an optimal signal for  $A$  if only marginal distributions of  $A$  and  $B$  were known.

Next we describe the simultaneous signal structure, i.e. when the sender decides to inform the receiver about both dimensions simultaneously.

### 2.3.4 Simultaneous signal structure and order of moves

Sender can decide to provide information simultaneously about both dimensions. This can be accomplished by making signal a family of conditional distributions on the joint state space and thus informing about both dimensions simultaneously. Now an element ( $s$ ) of the space of signal realisation ( $S$ ) can be regarded as a recommendation about two actions. Simultaneous signal is  $\pi(\cdot|\omega_{A,B}), \omega_{A,B} \in \{0, 1, \dots, n\} \times \{0, 1, \dots, n\}$ . Signal realisation is  $s_{(i,j)}$ , where,  $i$  and  $j \in \{0, \dots, n\}$ , which can be interpreted in the following way: choose  $i$  for dimension  $A$  and  $j$  for dimension  $B$ .

Simultaneous signal accomplishes the following: given signal realisation  $s_{(i,j)}$ , receiver forms beliefs about the joint distribution of  $A$  and  $B$ . Then, given new joint distribution, receiver calculates marginal distributions of  $A$  and  $B$  and makes decisions for both dimensions.

### 2.3.5 Optimal signals

We are interested in signals that maximise sender's expected payoff. Following Bergemann and Morris [2016a] sometimes we will refer to obedience constraint that signal should satisfy. This means for example that if signal realisation is  $s_{(i,j)}$ , it can be interpreted as a direct recommendation to the receiver to choose  $i$  for dimension  $A$  and  $j$  for dimension  $B$  and the receiver should have an incentive to follow this recommendation, i.e. it should be optimal for the receiver to choose  $i$  and  $j$ . Also sometimes, for easy of exposition, we will present the signal as a joint distribution of signal space and the decision relevant space.

## 2.4 General Observations

In the current section we want to relate two-dimensional bayesian persuasion of the current model to the one-dimensional bayesian problem. The latter is equivalent to the case when only marginal distributions of  $A$  and  $B$  are known, or when  $A$  and  $B$  are independent. First

we want to describe receiver's optimal payoff in terms of his payoff when only marginal distributions of  $A$  and  $B$  are known.

Preference specification of our model means that receiver's optimal decision about dimension  $i$  is only a function of expected state for  $i$ . This follows from the fact that receiver's preferences are additively separable across dimensions  $A$  and  $B$ . This means that additional information in the form of joint distribution of  $A$  and  $B$  acts like an additional constraint for the sender. We formalise this observation in the following proposition.

**Proposition 1.** *For arbitrary  $n$  and arbitrary joint distribution of  $A$  and  $B$  upper bound of sender's expected payoff is what the sender could get if only marginal distributions of  $A$  and  $B$  were known.*

*Proof.* Say there exists a signal that gives sender higher payoff than what he would get if only marginal distributions were known. This means that there exists a dimension  $i$  for which sender's payoff from the signal is higher than his optimal payoff for  $i$  if only marginal distribution of  $i$  was known. But because receiver's optimal action about dimension  $i$  is only a function of  $i$ 's expected state, this is a contradiction.  $\square$

This argument can be illustrated by the following reasoning. Say the receiver has to make a single decision about dimension  $i$ . Then the additional information about some irrelevant state of the world, in the current case in the form of joint distribution of  $i$  and the decision irrelevant state of the world, can be of no benefit for the sender. In the current model state of the world for dimension  $j$  is irrelevant when the receiver makes a decision about  $i$ .

After describing upper bound for sender's payoff, now we want to describe a necessary and sufficient condition for a sequential signal structure to achieve this upper bound. This will turn out helpful when finding optimal sequential signal structures.

If only marginal distributions of  $A$  and  $B$  were known, then we could calculate sender's optimal payoff by finding optimal signals for  $A$  and  $B$  separately, as suggested by Kamenica and Gentzkow [2011].

Note also that for these marginal distributions, if the common prior was joint distribution, such that  $A$  and  $B$  were independent, then optimal sequential signal structure would give sender the same payoff as when only marginal distributions are known. But if the common prior is a joint distribution of  $A$  and  $B$  and they are not independent, then the signal for one dimension in general will also contain an information about the other dimension. Before formulating our main argument, we make following two observations.

First, we have to remember that sender's value function,  $V$ , with signals, is concave by construction. This is the popular concavification argument as developed by Aumann and Maschler [1995] and Kamenica and Gentzkow [2011]. For exposition purposes we briefly repeat this argument. We follow here Kamenica and Gentzkow [2011].

Denote sender's value function by  $v(p)$ . Denote convex hull of  $v(p)$  by  $co(v)$ . Then, concave closure of  $v(p)$  is defined in the following way:

$$V(p) \equiv \{z | (p, z) \in co(v)\} \quad (2.1)$$

Kamenica and Gentzkow [2011] show that sender's optimal payoff for a prior  $p_0$  is  $V(p_0)$ .

We formalise these observations in the following lemma:

**Lemma 1.** *Sender's maximum payoff for belief  $p$  is  $V(p)$ . Sender's value function with signals,  $V$ , is concave by construction.*

Second observation is that the signal for  $i$  induces a probability distribution of posteriors of  $j$  that equals to the prior marginal distribution of  $j$ . This follows from the law of total probability. We formalise this observation in the following lemma:

**Lemma 2.** *Signal for  $i$  induces a distribution of  $j$ 's posteriors, whose expectation equals to  $j$ 's prior.*

Now we can prove the following result:

**Proposition 2.** *For arbitrary  $n$  and arbitrary joint distribution of  $A$  and  $B$ , there exists a sequential signal structure that achieves an upper bound, if and only if following is true:*

*There exists a dimension  $i$  s.t. first producing an optimal signal for  $i$  induces a distribution of posteriors of  $j$  whose support belongs to an interval on which sender's value function for  $j$  is linear.*

*Proof.* Say there exists such an  $i$ . Then, first producing an optimal signal for  $i$  gives following expected payoff for  $j$ :

$$\begin{aligned} p(s_{i0})V(\beta_{(s_{i0})}) + \dots + p(s_{in})V(\beta_{(s_{in})}) &= V(p(s_{i0})\beta_{(s_{i0})} + \dots + \\ &\quad p(s_{in})\beta_{(s_{in})}) = V(\beta) \end{aligned} \quad (2.2)$$

Where, for example  $p(s_{i0})$  is a probability of observing signal realisation  $s_{i0}$ , which leads to the posterior distribution  $(\beta_{(s_{i0})})$  of  $j$ .  $V(\beta_{(s_{i0})})$  is sender's payoff for dimension  $j$ , when he chooses an optimal signal for  $j$ , when the prior belief is  $(\beta_{(s_{i0})})$ .  $\beta$  is the prior marginal distribution of  $j$ .

First equality follows from our assumption that the value function of  $j$  is linear on the interval to which the support of the distribution of  $j$ 's posteriors belongs, as induced by the optimal signal of  $i$ . The second equality follows from lemma 2.

Now say there does not exist such an  $i$ . If the signal for the dimension  $i$  is optimal, then we get the following expression:

$$\begin{aligned} p(s_{i0})V(\beta_{(s_{i0})}) + \dots + p(s_{in})V(\beta_{(s_{in})}) &< V(p(s_{i0})\beta_{(s_{i0})} + \dots + \\ &\quad p(s_{in})\beta_{(s_{in})}) = V(\beta) \end{aligned} \quad (2.3)$$

Inequality follows from Jensen's inequality and lemma 1, because  $V$  is not linear on the support of the distribution of  $j$ 's posteriors as induced by the signal for  $i$ . Equality follows again from lemma 2.  $\square$

For illustration we will now go back to the example discussed in the introduction and give graphical arguments.



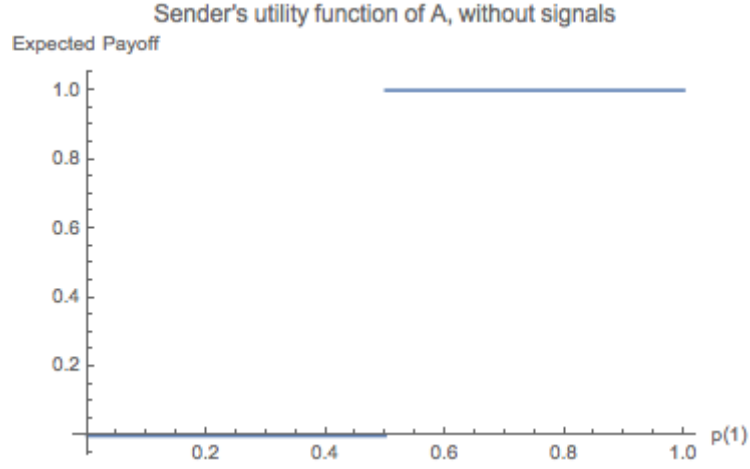


Figure 2.2

We want to show graphically how does the correlation between  $A$  and  $B$  affect sender's payoff for  $A$ , when he chooses to provide information sequentially and the first signal is produced for  $B$ . First we briefly describe how to construct sender's value function with signals.

One starts with sender's utility function without signals. Sender's utility without signals is 0 if  $p(A = 1) < \frac{1}{2}$  and 1 if  $p(A = 1) \geq \frac{1}{2}$  and is given in figure 2.2. Next we construct smallest concave function weakly bigger than sender's utility function. So, if only marginal distributions of  $A$  and  $B$  were known, then sender's maximum payoff for  $A$  in the example from the introduction would be 0.90, because smallest concave function everywhere weakly greater than sender's value function is  $\min\{2p(A = 1), 1\}$  and is given in figure 2.3.

When the first optimal signal is produced for  $B$ , then the posterior of  $A$  becomes bigger than 0.5 with positive probability. Therefore, the expected payoff for  $A$ , when the first signal is produced for  $B$ , is smaller than if only marginal distribution of  $A$  was known.

In the next section we analyse optimal simultaneous and sequential signal structures when  $n = 1$  for each dimension.

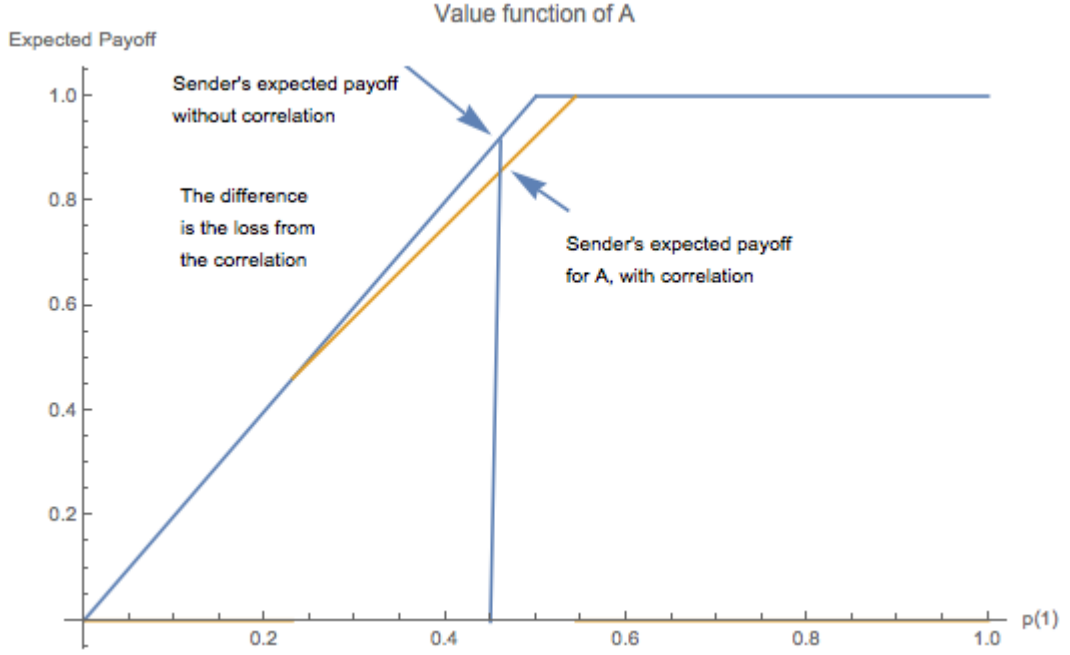


Figure 2.3: Concavification of value function; vertical line describes the set of feasible payoffs for  $A$ , when  $p(A = 1) = 0.45$  and  $A$  and  $B$  are independent.

## 2.5 Optimal simultaneous and sequential signal structures, when $n=1$

For the motivating example in the introduction, if the first signal is produced for  $A$  then there exists a sequential signal structure that achieves an upper bound. Following example 2.4<sup>1</sup> shows that this is not true in general, i.e. there exist joint distributions, for which there does not exist a sequential signal structure that achieves the upper bound.

To see this, note the following: if the first signal is produced for dimension  $i$ , then  $p(i = 0)$  is in the support of the distribution induced by the optimal signal for  $i$ . But  $p(j = 1|i = 0) = \frac{3}{5}$ . Therefore, it follows from proposition 2 that there does not exist a sequential signal structure that achieves the upper bound.

This example shows that there exist joint distributions of  $A$  and  $B$ , for which whatever the order of persuasion, if the first signal pro-

<sup>1</sup>This example was suggested by Sergiu Hart.

	$B = 1$	$B = 0$
$A = 1$	$\frac{1}{9}$	$\frac{3}{9}$
$A = 0$	$\frac{3}{9}$	$\frac{2}{9}$

Figure 2.4: Example, where no sequential signal can achieve an upper bound

duced is optimal, than the sequential information provision always reveals too much information to the receiver, than what is optimal for the sender.

So, the problem with a sequential signal structure is that it can reveal too much information to the receiver. To hinder the receiver to learn about one dimension from the signal about the other dimension signal should inform about the joint state.

Now we derive an optimal simultaneous signal that conditions on the joint states and therefore induces a distribution of posterior beliefs about joint states.

### 2.5.1 Optimal simultaneous signal

To calculate optimal simultaneous signal when there are four states, we can not use concavification of the value function, since we will require four dimensions to visualise sender's value function as a function of distribution of beliefs. Instead, following Bergemann and Morris [2016a] we will think about signal as a recommendation rule for the receiver, that should satisfy obedience constraint. For example, if signal realisation is  $s_{(1,1)}$ , then receiver would choose 1 for A and B iff  $p(i = 1|s_{(1,1)}) \geq \frac{1}{2}$ , for  $i \in \{A, B\}$ . For ease of notation we will represent the signal as a joint distribution of decision relevant state space and signal space, i.e. joint distribution of  $\omega_{A,B}$  and  $s_{(i,j)}$ . Let's denote joint distribution of A and B in the following way:

	$B$	
$A$	$p(1,1)$	$p(1,0)$
	$p(0,1)$	$p(0,0)$

where, for example,  $p(1,0)$  denotes probability of the following event:  $A = 1$  and  $B = 0$ .

We will also use the following notation. Consider states  $(1,0)$  and  $(0,1)$ : let  $\alpha$  denote the state that is less likely among these states and  $\beta$  the state that is more likely. For example if  $p(1,0) > p(0,1)$ , then  $\alpha = (0,1)$  and  $\beta = (1,0)$ . Also, if  $p(1,0) = p(0,1)$ , then say  $\alpha = (1,0)$  and  $\beta = (0,1)$ . Then expression  $p(\beta)$  would mean more likely event among  $\{(1,0), (0,1)\}$ , if  $p(1,0) \neq p(0,1)$  and  $p(0,1)$  otherwise. Also we will use the above notation of states for recommendations, i.e.  $s_{(\alpha)}$  and  $s_{(\beta)}$ .

Now we can proof the following result:

**Proposition 3.** *A optimal simultaneous signal is given by the following joint distribution:*

$s/\omega$	$(1,1)$	$\alpha$	$\beta$	$(0,0)$
$s_{(1,1)}$	$p(1,1)$	$p(\alpha)$	$p(\alpha)$	$p(1,1)$
$s_{(\alpha)}$	0	0	0	0
$s_{(\beta)}$	0	0	$p(\beta) - p(\alpha)$	$p(\beta) - p(\alpha)$
$s_{(0,0)}$	0	0	0	$p(0,0) - [p(1,1) + p(\beta) - p(\alpha)]$

*Proof.* First, note that signal recommendations satisfy obedience constraints. Second, note that the payoff from this signal is the same as what would be if only marginal distributions of  $A$  and  $B$  were known, as induced by the joint distribution. Then, it follows from the proposition 1 that this signal is optimal. □

The suggested optimal simultaneous signal is not unique. Note that the signal does not make a recommendation of  $\alpha$  if  $p(\alpha) < p(\beta)$ , i.e.  $p(s_\alpha) = 0$ ; and if  $p(\alpha) = p(\beta)$ , then  $p(s_\alpha) = p(s_\beta)$ . One can relatively easily construct optimal signals where all four states are recommended with positive probability. But for all these signals following is true: support of the distribution of posterior joint distributions is such that marginal distributions as induced by these joint distributions always belong to the support of the distribution of marginal distributions as induced by optimal signals when only marginal distributions

of  $A$  and  $B$  are known. To put it simply, we know that when only marginal distributions are known, then the support of the distribution of posteriors as induced by the optimal signal is 0 and  $\frac{1}{2}$ . This is true also for the suggested optimal simultaneous signal. What we are saying is that although the signal is not unique, this property remains true for other optimal simultaneous signals as considered here. It still remains to be shown formally that this claim is true in general, i.e. for all optimal simultaneous signals.

After deriving sender's optimal simultaneous signal, we want to analyse optimal sequential signal structures.

## 2.5.2 Sequential signal structure

Our goal is to find optimal sequential signal structures, when  $n = 1$ .

As we showed above, sender's value function with signals for dimension  $i$  is the following:

$$V(p(1)) = \min\{2p(1), 1\} \quad (2.4)$$

From equation (2.4) follows that sender's value function with signals is linear on the interval  $[0, 0.5]$ . Therefore if the first signal is produced for  $A$  then for a sequential signal structure to achieve the same payoff as when only marginal distributions are known, it has to be the case that support of the distribution of  $B$ 's posteriors, as induced by the optimal signal for  $A$ , has to be in the interval  $[0, 0.5]$ . This follows from proposition 2.

**Corollary 1.** *There exists a sequential signal structure that achieves the same expected payoff for the sender as what he optimally can get if only marginal distributions were known, if and only if following is true: there exists a pair  $(i, j)$  s.t. first producing an optimal signal for  $i$  induces a distribution of  $j$ 's posteriors with support  $[0, 0.5]$ .*

We have seen that our motivating example 2.1 allowed sequential signal structure that achieved an upper bound for the sender, whereas our second example 2.4 showed that there are joint distributions, for

which no sequential signal can achieve an upper bound. One difference between these examples is that in the first example  $A$  and  $B$  are positively correlated, whereas in the second case they are negatively correlated.

Now we want to characterise joint distributions in terms of correlation that allow sequential signal that achieves an upper bound.

Signal for  $i$  directly informs only about  $i$ . Thus, optimal signal for  $i$  induces two posteriors of  $j$ , one for  $s_{i1}$  and another for  $s_{i0}$ . It follows from the concavification argument that optimal signal for  $i$  induces a distribution of posterior beliefs of  $i$ , that has following support:  $p(i = 1|s_{i1}) = \frac{1}{2}$  and  $p(i = 1|s_{i0}) = 0$ . Now we are interested in the support of  $j$ 's posterior beliefs, as induced by the optimal signal for  $i$ . This distribution has following support, expressed in terms of conditional probabilities:

$$p(j = 1|s_{i1}) = \frac{1}{2} [p(j = 1|i = 1) + p(j = 1|i = 0)] \quad (2.5)$$

$$p(j = 1|s_{i0}) = (j = 1|i = 0) \quad (2.6)$$

In the current discussion  $s_i$  denotes an element of an optimal signal for  $i$ , i.e. signal that would be optimal if only marginal distributions of  $A$  and  $B$  were known.

It turns out that if  $A$  and  $B$  are positively correlated then there always exists a sequential signal structure that achieves an upper bound; while in the case of negative correlation we characterise a sufficient and necessary condition for a sequential signal structure to achieve an upper bound. Before proving these claims, we need to show some preliminary results.

First note that since we are considering the case when there is a gain from persuasion in both dimensions, from this trivially follows that  $p(j = 1|s_{i0}) > \frac{1}{2}$  only if  $A$  and  $B$  are negatively correlated.

First we show that if  $p(j = 1|s_{i1}) > \frac{1}{2}$  then  $p(i = 1|s_{j1}) < \frac{1}{2}$ .

**Lemma 3.** *If  $p(j = 1|s_{i1}) > \frac{1}{2}$ , then  $p(i = 1|s_{j1}) < \frac{1}{2}$ .*

*Proof.* Say  $p(A = 1|s_{B1}) > \frac{1}{2}$ . By remembering that  $p(B = 1|s_{B1}) = \frac{1}{2}$ , then one has the following:

$$p(A = 1|s_{B1}) = p(A = 1|B = 1)\frac{1}{2} + p(A = 1|B = 0)\frac{1}{2} > \frac{1}{2} \quad (2.7)$$

After substituting expressions for conditional probabilities and simplifying, one sees that inequality (2.7) holds iff following is true:

$$p(1,1)p(1,0) > p(0,0)p(0,1). \quad (2.8)$$

Note that  $p(1,1) < p(0,0)$ , otherwise for at least one dimension  $p(i = 1) > \frac{1}{2}$ . Therefore  $p(1,0) > p(0,1)$ .

Say now following is also true:

$$p(B = 1|s_{A1}) > \frac{1}{2} \quad (2.9)$$

By the same argument as above, one can show that inequality (2.9) holds iff

$$p(1,1)p(0,1) > p(0,0)p(1,0) \quad (2.10)$$

which is a contradiction. □

It is also straightforward that if A and B are negatively correlated then  $p(j = 1|s_{i1}) < \frac{1}{2}$ . We formalise this in the next lemma.

**Lemma 4.** *If A and B are negatively correlated then  $p(j = 1|s_{i1}) < \frac{1}{2}$ .*

*Proof.* For concreteness consider the case when the first signal is produced for A. We want to show that if A and B are negatively correlated then  $p(B = 1|s_{A1}) < \frac{1}{2}$ . First note that negative correlation, when there are two states for each dimension, implies the following:  $p(i = 1|j = 1) < p(i = 1) < p(i = 1|j = 0)$ . By using the law of total probability, we can express probability of  $B = 1$  in the following way:

$$p(B = 1) = p(B = 1|A = 1)p(A = 1) + p(B = 1|A = 0)p(A = 0).$$

The result follows from noting that  $p(A = 1|s_{A1}) > p(A = 1)$  and  $p(B = 1) < \frac{1}{2}$ .

□

Now we can prove the following result:

**Proposition 4.** (a) *If A and B are positively correlated then there exists a sequential signal structure that achieves an upper bound.*

(b) *If A and B are negatively correlated then there exists a sequential signal structure that achieves an upper bound iff following is true: there exists a pair i and j, s.t. :  $p(j = 1|i = 0) \leq \frac{1}{2}$ .*

*Proof.* (a) Say A and B are positively correlated. Then we know from lemma (3) together with the fact that  $p(j = 1|i = 0) < \frac{1}{2}$  is always true for positive correlation, that there exists a random variable j, s.t support of the distribution of it's posteriors as induced by the optimal signal of i is always in the interval  $[0, 0.5]$ . The result then follows from corollary (1).

(b) Say there exists a pair i and j s.t.  $p(j = 1|i = 0) \leq \frac{1}{2}$ . Then it follows from lemma (4) that there exists a random variable j s.t. support of it's distribution as induced by the optimal signal of i belongs to the interval  $[0, 0.5]$ . The result then follows from corollary (1). Say there does not exist a pair for which  $p(j = 1|i = 0) \leq \frac{1}{2}$  is true. Then it follows again from corollary (1) that there does not exist a sequential signal structure that achieves an upper bound. □

Based on the previous results one can also give a simple characterisation of sequential signal structures that achieve an upper bound.

**Corollary 2.** *If A and B are positively correlated, then following sequential signal structure achieves an upper bound: first produce an optimal signal for the dimension, whose expectation is not smaller than the expectation of the other dimension. Given new posteriors induced by this signal, then produce an optimal signal for another dimension.*

**Corollary 3.** *If A and B are negatively correlated and there exists a pair, s.t.  $p(j = 1|i = 0) \leq 0$ , then following sequential signal structure achieves*



*an upper bound: first produce an optimal signal for  $i$  and then given new posteriors of  $j$ , produce an optimal signal for  $j$ .*

As we have seen in the example 2.4, when A and B are negatively correlated, then there exists a joint distribution for which there is no sequential signal structure that achieves an upper bound. So, question remains what is an optimal sequential signal in this case, i.e. when following is true:

$$p(A = 1|B = 0) > \frac{1}{2} \quad (2.11)$$

$$p(B = 1|A = 0) > \frac{1}{2} \quad (2.12)$$

The intuition is that the inefficiency increases in the distance  $p(j = 1|i = 0) - \frac{1}{2}$ . It turns out that the intuition is correct and it is optimal first signal to produce for  $i$ , for which following is true:

$$p(i = 1|j = 0) \geq p(j = 1|i = 0) \quad (2.13)$$

Before proving this result, we want to make some observations, that will turn out helpful. First note that it can never be optimal for the sender to not persuade receiver about some dimension. We formulate this observation in the following lemma, but first we give a definition of an uninformative signal.

**Definition 1.** *A signal is uninformative, if cardinality of the set of signal realisations is singleton. A signal is informative, if it is not uninformative.*

**Lemma 5.** *For sender it is never optimal to not produce informative signal about some dimension.*

*Proof.* Say sender decides to produce uninformative signal about dimension A. Then, because order of persuasion is not given, sender can decide to produce first signal about B. Note that for any signal of B, posterior of A will be smaller than 0.5 with positive probability. Then sender strictly gains from producing signal for A.  $\square$

Now we want to argue that whatever the dimension, for which the first signal is produced, this signal should be the same that the sender would choose if only marginal distributions of  $A$  and  $B$  were known.

**Lemma 6.** *Say the sender decides to produce the first signal about dimension  $i$ . Then the sequential signal structure is optimal only if the signal for  $i$  is the same that the sender would choose if only marginal distributions of  $A$  and  $B$  were known.*

*Proof.* From lemma 5 follows that whatever the dimension for which the first signal is produced, posterior for this dimension should become at least 0.5 with positive probability. For simplicity we assume that the first signal is produced for  $A$ . It is not difficult to see that it can never be optimal to produce a signal for which posterior of  $A$  becomes bigger than 0.5. Examining expected payoff for  $B$  graphically should be enough to see that this claim is correct, by remembering that  $A$  and  $B$  are negatively correlated.

Another possibility could be that the signal for  $A$  is such that the posterior of  $B$  never becomes bigger than 0.5, so that the expected payoff from  $B$  is the same as when only marginal distributions are known. We will show that this can not be optimal.

First we want to calculate sender's optimal expected payoff when the signal for  $A$  is such that the support of the distribution of  $B$ 's posteriors is  $[0, 0.5]$ . Optimisation implies following: we want to find signal for  $A$ , that has following properties:  $p(A = 1|s_{A1}) = \frac{1}{2}$  and  $p(A = 1|s_{A0})$  is such that  $p(B = 1|p(A = 1|s_{A0})) = \frac{1}{2}$ . Solving for optimal signal of  $A$  under this constraint gives the following expected payoff:

$$\frac{(p(B = 1|A = 0) - p(B = 1|A = 1))p(A = 1) - p(B = 1|A = 0) + 0.5}{0.5 - 0.5(p(B = 1|A = 0) + p(B = 1|A = 1))} + 2p(B = 1) \quad (2.14)$$

Now we want to calculate sender's payoff from a sequential signal structure, when the first signal is produced for  $A$  and it is the same

signal that would be optimal if only marginal distribution of  $A$  was known. This is given by the following expression:

$$1 + 2p(A = 1) [p(B = 1|A = 1) + p(B = 1|A = 0)] \quad (2.15)$$

Now, after substituting expressions for conditional distributions and expressing marginal distributions in terms of joint distribution, it turns out that 2.14 can never be bigger than 2.15, when  $A$  and  $B$  are negatively correlated and prior expectation for each dimension is smaller than 0.5.

This completes the proof. □

Now we can prove the following result:

**Proposition 5.** *If  $A$  and  $B$  are negatively correlated and if a sequential signal is chosen, s.t. first signal is produced for  $i$  for which following is true:  $p(i = 1|j = 0) \geq p(j = 1|i = 0)$ , then there does not exist a sequential signal structure that achieves higher expected payoff for the sender.*

*Proof.* If  $p(i = 1|j = 0) \leq \frac{1}{2}$ , then we showed above that such a signal achieves an upper bound.

Say now  $p(j = 1|i = 0) > \frac{1}{2}$ , so that no sequential signal achieves an upper bound. It follows from lemmas 5 and 6 that whatever the dimension for which the first signal is produced, this signal should be the same as when only marginal distributions of  $A$  and  $B$  are known. Now it remains to compare which of the two sequential signals give higher payoff.

First producing an optimal signal for  $A$  gives following expected payoff to the sender:

$$1 + 2[p(1, 1) + p(1, 0)] \left[ \frac{p(1, 1)}{p(1, 1) + p(1, 0)} + \frac{p(0, 1)}{p(0, 1) + p(0, 0)} \right] \quad (2.16)$$

Sender's expected payoff from producing first signal for  $B$  is:

$$1 + 2[p(1,1) + p(0,1)]\left[\frac{p(1,1)}{p(1,1) + p(0,1)} + \frac{p(1,0)}{p(1,0) + p(0,0)}\right] \quad (2.17)$$

After subtracting expression 2.17 from expression 2.16, one gets:

$$[p(0,1) - p(1,0)][p(1,1)p(0,0) - p(1,0)p(0,1)] \quad (2.18)$$

First note that the negative correlation implies that the second term is negative. Regarding the sign of the first term, one has the following: if  $p(B = 1|A = 0) > (<)p(A = 1|B = 0)$  then the first term is positive (negative). This ends the proof.  $\square$

Now we have completely analysed optimal simultaneous and sequential signal structures when  $n = 1$  for each dimension. We derived an optimal simultaneous signal that always achieves an upper bound of sender's payoff. Then we fully characterised optimal sequential signal structures. We characterised joint distributions for which sender gets the same payoff as what he would get from the simultaneous signal. So for this class of joint distributions the problem allows simple approach, since one can consider signals on the smaller state space.

Now, after having fully analysed the binary case, we want to understand when does a sequential signal achieve the same payoff as when only marginal distributions are known for the case when  $n = 2$  for each dimension.

## 2.6 Characterising conditions for a sequential signal structure to achieve an upper bound when $n=2$

Our goal is again to characterise conditions for a sequential signal structure to achieve an upper bound. First step is to characterise optimal signals and value function with 3 states for one dimension only. The derivation of optimal signals we relegate to the appendix and here

give sender's value function.

**Lemma 7.** *When  $n = 2$ , receiver's value function for one dimensional persuasion problem is the following:*

$$2(p(1) + 2p(2)) \quad \text{if } p(1) + 2p(2) < 0.5 \quad (2.19)$$

$$2p(2) + p(1) + 0.5 \quad \text{if } 0.5 \leq p(1) + 2p(2) < 1.5 \text{ and } p(1) \leq 0.5 \quad (2.20)$$

$$2p(2) + 1 \quad \text{if } 0.5 \leq p(1) + 2p(2) < 1.5 \text{ and } p(1) > 0.5 \quad (2.21)$$

$$2 \quad \text{if } p(1) + 2p(2) \geq 1.5 \quad (2.22)$$

Now we are ready to characterise conditions for the existence of a sequential signal structure to achieve an upper bound. Value function is linear function on three different intervals and constant if the expected state is at least 1.5. Therefore for a sequential signal structure to achieve an upper bound it must be the case that the posteriors of  $B$ , if the first signal is produced for  $A$ , should remain in the interval, to which the prior belongs. We formalise this in the next corollary.

**Corollary 4.** *There exists a sequential signal structure that achieves an upper bound, iff following is true: there exists a pair  $i$  and  $j$ , s.t. first producing an optimal signal for  $i$ , the support of the distribution of posteriors of  $j$  belongs to the same interval to which the prior of  $j$  belonged.*

*Proof.* This follows from proposition 2. □

## Appendix 2.A Optimal signals when $n=2$ and only one decision is to be made

First we describe receiver's best action as a function of his beliefs. If all three states are possible, then receiver's optimal action is:

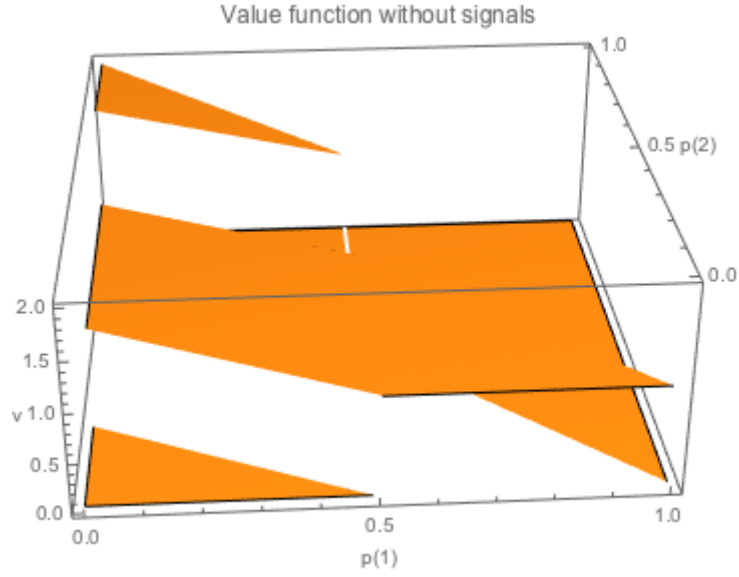


Figure 2.5

$$\begin{aligned}
 &0 \quad \text{if} \quad 2p(2) + p(1) < \frac{1}{2} \\
 &1 \quad \text{if} \quad \frac{1}{2} \leq 2p(2) + p(1) < \frac{3}{2} \\
 &2 \quad \text{if} \quad 2p(2) + p(1) \geq \frac{3}{2}
 \end{aligned}$$

Receiver chooses action closest to the expected state. To derive sender's value function and optimal posterior distributions we will construct sender's value function without signal and from this we get the value function with signals by concavification. Although receiver is only interested in the expected state, the value function of the sender is a function of the distribution of states and not an expectation. Kamenica and Gentzkow [2011] discuss this question in some detail. Therefore we have to analyse the value function of the sender as a function of two probabilities,  $p(1)$  and  $p(2)$ . The value function without signals is a step function with values 0, 1 and 2 and is given by figure 2.5. By concavification of this function one gets sender's value function with signals, as depicted in figure 2.6.

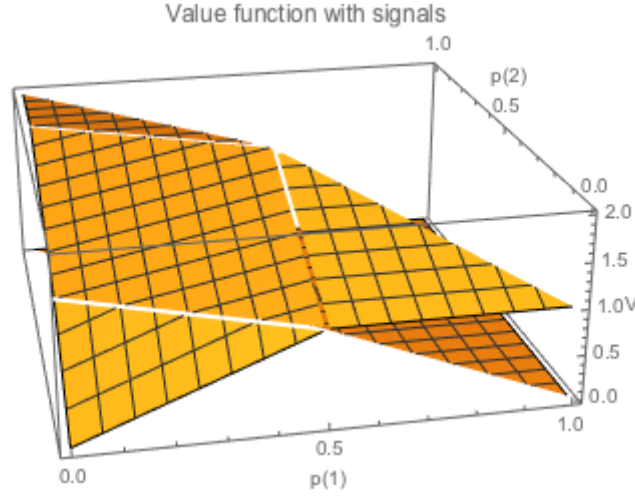


Figure 2.6

We derive now optimal distribution of posterior beliefs, s.t. expectation equals prior. This is equivalent to deriving optimal signals. To derive an analytic expression of the value function it will be helpful to look at the domain of the value function. Figure 2.7 gives the level-areas of different values of the function. First let's consider the area labeled by (a). These are the combinations of  $p(1)$  and  $p(2)$ , for which the expected value is smaller than 0.5 and therefore receiver's optimal action is 0. One can see that the optimal signal in this case means choosing distribution of posteriors s.t. its expectation equals to the prior and support of this distribution is  $2p(2) + p(1) = \frac{1}{2}$  and  $p(2) = p(1) = 0$ . Solution of this problem is the following:

$$p(s_1) = 2(p(1) + 2p(2)) \quad (2.23)$$

$$p(s_2) = 0 \quad (2.24)$$

$$p(1|s_1) = \frac{p(1)}{2(p(1) + 2p(2))} \quad (2.25)$$

$$p(2|s_1) = \frac{p(2)}{2(p(1) + 2p(2))} \quad (2.26)$$

$$p(1|s_0) = 0 \quad (2.27)$$

$$p(2|s_0) = 0 \quad (2.28)$$

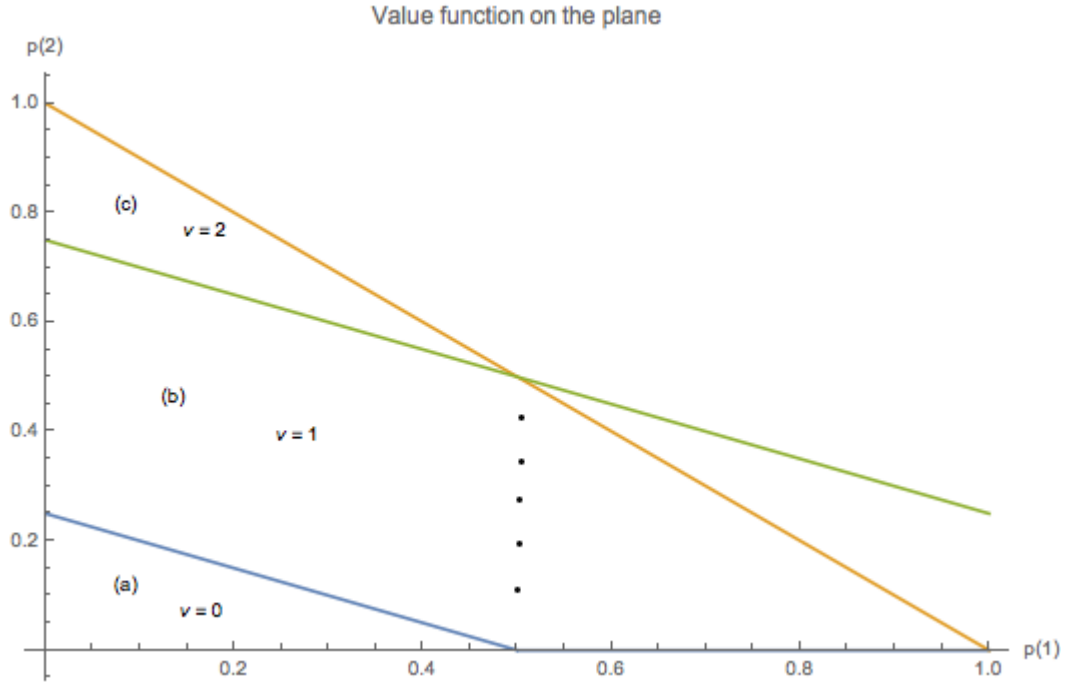


Figure 2.7

Now we want to find optimal distribution of posteriors for the priors, for which the best action of the receiver is 1. The distribution of priors for which receiver's optimal action is 1 is labeled by (b). For these distributions of priors the goal is to induce distributions of posteriors which make actions 2 and 1 optimal for the receiver. Here one has to distinguish between two cases:  $p(1) < 0.5$  and  $p(1) > 0.5$ . When  $p(1) < 0.5$ , then solution of this problem is the following:

$$p(s_1) = 1.5 - p(1) - 2p(2) \quad (2.29)$$

$$p(s_2) = p(1) + 2p(2) - 0.5 \quad (2.30)$$

$$p(1|s_1) = p(1) \quad (2.31)$$

$$p(1|s_2) = p(1) \quad (2.32)$$

$$p(2|s_2) = 0.75 - 0.5p(1) \quad (2.33)$$

$$p(2|s_1) = 0.25 - 0.5p(1) \quad (2.34)$$

If  $p(1) > 0.5$ , then solution is the following:



$$p(s_1) = 1 - 2p(2) \quad (2.35)$$

$$p(s_2) = 2p(2) \quad (2.36)$$

$$p(1|s_2) = 0.5 \quad (2.37)$$

$$p(2|s_2) = 0.5 \quad (2.38)$$

$$p(1|s_1) = 0.5 \frac{p(1) - p(2)}{0.5 - p(2)} \quad (2.39)$$

$$p(2|s_1) = 0 \quad (2.40)$$

## Chapter 3

# Identifying the reasons for coordination failure in a laboratory experiment

### 3.1 Introduction

If you have lost your spouse in a department store and both of you are trying to find each other, the answer to the (seemingly simple) question of “Will she look for me at the coffee bar or at the exit?” depends not only on the answer to the question “Does she think I am looking for her at the coffee bar or at the exit?” (i.e., something we will call the first-order belief) but also on the answers to “Does she think that I think that she thinks that I am looking for her at the coffee bar or at the exit?” (i.e., the second-order belief or “What is her first-order belief?”) and on infinitely more levels of beliefs. This paper addresses the question if people actually use beliefs of a higher order.

When modeling human behavior, most works assume that players have common knowledge about the structure of the game, i.e., that all players know the structure, that all players know that everyone else knows the structure and so on.<sup>1</sup> Furthermore, we assume that players

---

<sup>1</sup>However, there is a (mostly theoretical) literature on universal type spaces, introduced by Harsanyi [1967/68] and formalized by Mertens and Zamir [1985] and Brandenburger and Dekel [1993], analysis games without common knowledge.

do not only have common knowledge about publicly known properties of the game but also about the distributions of unknown factors of the game, like the other players' types (for example if I'd rather wait at the coffee bar or the exit). The absence of common knowledge leads to complex belief hierarchies, so called *higher-order beliefs*. The first level of these beliefs, so called first-order beliefs, might be a belief over the other player's type. A second-order belief would then be a belief over the belief of the other player about your type (i.e., a belief over the other player's first-order belief) and so on ad infinitum.

In the game theoretical literature many different assumptions and models of higher-order beliefs exist and many of these lead to very different predictions even in simple games like the pure coordination game we are using in this paper.<sup>2</sup> The question, what kind of model of higher-order beliefs players actually use, seems to be an empirical question which we are trying to address in this paper.

To do so, we take up the experimental results and setup of Blume and Gneezy [2010], in which there is an issue of cognitive difficulties, to analyze the effects of higher-order beliefs. Blume and Gneezy [2010] used a slightly difficult coordination game (the so-called 5-sector disc), in which there is a better (i.e., risk- and payoff-dominant) option which is harder to find. They were able to show that participants form beliefs about the cognitive abilities of other participants and, if these beliefs are pessimistic, they hinder coordination between the players (i.e., that "beliefs matter"). However, they have not taken into account the effect of higher-order beliefs about cognitive abilities. Therefore, we modify their experimental setup in order to distinguish the effect of first-order beliefs players form about the cognitive ability of their opponents (i.e., if players trust in the cognitive ability of their partners) and higher-order beliefs.

We introduce a new treatment in which participants guess what other participants play against themselves. This allows us to identify first-order beliefs and therefore separate first- from higher-order be-

---

<sup>2</sup>A brief overview of some models of higher-order beliefs can be found in Section 3.1.1 and a more detailed discussion in Section 3.B.

liefs.

Using the data from these treatments, we can answer the following three questions:

- **Are players able to coordinate in the absence of common knowledge?**
- Can coordination fail because players underestimate the skill of the other players? Or, in other words, **do first-order beliefs matter?**
- Can coordination fail because players think "too much" about what others might think? Or, in other words, **do higher-order beliefs matter?**

Using Blume and Gneezy's [2000] 5-sector disc, we were able to find answers to all three questions: In the experiment, we were able to reproduce Blume and Gneezy's [2010] result, that the majority of players had no problem choosing the Pareto-dominant equilibrium strategy of the game (i.e., coordination is possible). Furthermore, some players switch to the worse equilibrium strategy because of first- and higher-order beliefs (i.e. first- and higher-order beliefs matter).

More important applications than the search for ones husband or wife in a department store are suggested by recent studies in sociology and development studies, like Bicchieri [2005]. She suggests that common knowledge plays a significant role in the fight against female genital mutilation. Our results might help to improve our understanding of why some organizations are significantly more successful in the fight against female genital mutilation (FGM) than others. This application is discussed in more detail in Section 3.6.

Apart from the FGM application, higher-order beliefs have applications in many different fields. For example, they might be a reason for bank runs (i.e., the belief that a bank will be fine but a fear that others might think that the bank is in trouble might cause a bank run on this bank), arms races and financial crises.

The paper is organized as follows: In Section 3.1.1, we will give an overview of the relevant literature and how our work fits into it. Then we will explain an example of the game we use in Section 3.2.

In Section 3.3 we will explain the model. This is followed by the experimental design in Section 3.4 and the results of the experiment in Section 3.5. The aforementioned application to the fight against female genital mutilation is discussed in Section 3.6. Finally, we will conclude in Section 3.7.

### 3.1.1 Related works

There is a large theoretical literature, beginning with the seminal paper on the “email game” by Rubinstein [1989], showing that higher-order beliefs play a role in determining the outcome of a game. For instance, Carlsson and Van Damme [1993] use higher-order beliefs (in their model of global games) to identify the risk-dominant equilibrium as the unique rationalizable outcome of the coordination game. This uniqueness result spawned a large applied literature on, among other areas, bank runs and arms races, in e.g. Morris and Shin [1998], Morris and Shin [2004], Baliga and Sjöström [2004], Corsetti, Dasgupta, Morris, and Shin [2004], and Goldstein and Pauzner [2005]. Weinstein and Yildiz [2007b], however, have shown that this uniqueness result, that this whole literature depends on, is fragile to the exact specification of the higher-order belief model. Other “nearby” higher-order belief models have very different “unique” predictions. In fact, they show that any rationalizable outcome of the original game, can be obtained as the unique rationalizable strategy profile of some higher-order belief model.

Weinstein and Yildiz [2007a] establish a condition, called “global stability under uncertainty”. This condition implies that, if the change in equilibrium actions is small in the change of  $k$ th-order beliefs and higher, equilibria can be approximated by the equilibrium with at most  $k$ th-order beliefs. Unfortunately, pure coordination games do not fulfill “global stability under uncertainty”.

Strzalecki [2014] and Kneeland [2016] develop different non-equilibrium approaches, inspired by the experimental literature discussed later, using bounded levels of reasoning to explain behavior in coordinated at-

tack problems (e.g. Rubinstein’s [1989] email game).

A more in-depth discussion of models of higher-order beliefs and their predictions of the results of our experiment can be found in Section 3.B.

The experimental literature, however, has so far mostly focused on strategic uncertainty. The most prominent example for this is probably the literature on level-k thinking or cognitive hierarchy models, which was started by Nagel [1995] and Stahl and Wilson [1995]. In recent years, there have been many studies conducted, using and analyzing level-k reasoning, for example Ho, Camerer, and Weigelt [1998], Costa-Gomes, Crawford, and Broseta [2001], Camerer, Ho, and Chong [2004] and Crawford, Gneezy, and Rottenstreich [2008]. For a recent survey, see Crawford, Costa-Gomes, and Iriberri [2013].

But there also have been works which do not focus on strategic uncertainty. For example Heinemann, Nagel, and Ockenfels [2004], Cornand [2006], Cabrales, Nagel, and Armenter [2007] and Duffy and Ochs [2012] who directly test implications of the theory of global games, i.e. individuals play an incomplete information game as in Carlsson and Van Damme [1993]. The results however, are mixed and range from full support to full rejection of the predictions made by global games.

Another, closely related work is Kneeland [2015], in which she explores the level of rationality, a requirement for higher-order beliefs, of players experimentally. She shows that, in her experiment, 94% of all players are rational with decreasing numbers for second- (71%), third- (44%) and forth-order (22%) rationality.

We explore experimentally the “depth of reasoning” individuals employ when playing slightly difficult coordination games. In fact we want to abstract away from purely strategic concerns by only looking at coordination games in which the incentives of the players are perfectly aligned and a Pareto-dominant equilibrium exists. The fundamental uncertainty in the model will be one about the cognitive abilities of the opponents.

Differences in cognitive abilities have been studied before, for example by Gill and Prowse [2016], who have shown that more cogni-

tively able subjects converge, in repeated p-beauty contests, more frequently to equilibrium play and earn more. Furthermore, Proto, Rustichini, and Sofianos [2014] have shown that intelligence affects the results of repeatedly played prisoner's dilemmas, in which groups of higher intelligence tend to cooperate more frequently in later stages of the game. Agranov, Potamites, Schotter, and Tergiman [2012] have shown, by manipulating the perception of the cognitive levels of other players, that beliefs about the level of reasoning do play a significant role in the presence of strategic uncertainty. Alaoui and Penta [2015] establish a framework in which the depth of reasoning is endogenously determined by different cognitive costs of reasoning.

The way we model cognitive differences however, builds on another branch of literature. Motivated by Schelling's [1960] discussion of focal points, a variety of authors have tried to formally capture his ideas, most notably Bacharach [1993] and Sugden [1995]. The importance of focal points is supported by many experiments, for example by Mehta, Starmer, and Sugden [1994], who have replicated Schelling's results and have shown that coordinating on a focal point is different from accidental coordination. Crawford, Gneezy, and Rottenstreich [2008] have shown that, in a pure coordination game with symmetric payoffs, salient labels lead to a high percentage of coordination whereas even slight asymmetries in payoffs might lead to a coordination failure. Isoni, Poulsen, Sugden, and Tsutsui [2013] extend the analysis to bargaining problems and show that payoff-irrelevant clues help to improve coordination, even if there is no efficient or equal division.

In the absence of clues however, the theory of focal points can not be applied. Formally the absence of clues can be modeled as symmetries between strategies and players in a given game. In fact Nash [1951] has already discussed equilibrium under symmetry restrictions (and shown existence also of such symmetric (mixed) equilibria for finite games). Crawford and Haller [1990] have defined symmetries in games and used these definitions to see what focal points in highly

symmetric repeated coordination games would look like.<sup>3</sup> Blume [2000] has further developed this symmetry concept to talk about play under the absence of a common language. Other notions of symmetries have been put forward and studied in Harsanyi and Selten [1988], Casajus [2000] and Casajus [2001]. Alós-Ferrer and Kuzmics [2013] have then clarified the difference between different notions of symmetries and characterized all the possible ways a frame (the way a game is presented to players in the lab, for instance) could lead to different symmetry restrictions (and therefore to different focal points).

All these models of symmetries and restrictions are implicitly or explicitly investigated under the assumption of perfectly rational individuals. However, identifying all symmetries (and especially non-symmetries) in a game can be a difficult task. Bacharach [1993] has proposed his variable frame theory to allow for individual players with different states of mind or, as developed by Blume [2000] and employed by Blume and Gneezy [2000] and Blume and Gneezy [2010], with different cognitive abilities.

This finally brings us to the goal of our study. We want to take up the experimental results and setup of Blume and Gneezy [2010] to analyze the effects of higher-order beliefs. They were able to show that participants form beliefs about the cognitive abilities of other participants and, if these beliefs are pessimistic, they hinder coordination between the players. However, they have not taken into account the effect of higher-order beliefs about cognitive abilities. Therefore, we modify their experimental setup in order to distinguish the effect of first-order beliefs players form about the cognitive ability of their opponents and higher-order beliefs.

---

<sup>3</sup>Bhaskar [2000] and more comprehensively Kuzmics, Palfrey, and Rogers [2014], have studied theoretically and in the latter case also experimentally, what the possible focal points of the symmetric repeated battle-of-the-sexes and its generalizations could be.



### 3.2 Example

In this example, players only have access to two strategies  $l$  and  $h$  and are trying to coordinate on one of them; the payoffs are as depicted in the payoff matrix in Figure 3.1. As  $(h, h)$  has a higher equilibrium payment it would therefore be the focal point (and the risk- and payoff-dominant Nash equilibrium) of this particular game.<sup>4</sup>

	l	h
l	1,1	0,0
h	0,0	3,3

Figure 3.1: Payoff matrix of a high-cognition player

However, if we introduce cognitive differences, i.e., if action  $h$  is only available to a high-cognition player and low-cognition players are not aware of the existence of action  $h$  (and the high type) and are therefore forced to play  $l$ , beliefs about the other player's type might lead to coordination failure,<sup>5</sup> even if both players are high-cognition players. The driving force of this result is the absence of common knowledge about the players' type or the fraction of high cognition players.

This game models the situation in which one player is not aware that there even is an action to take (i.e., they don't have complete knowledge about the structure of the game).

The following two examples show how beliefs could lead to coordination failure between two high-cognition players: First imagine that the first player (she) thinks that the other player (he) is a low-cognition player. Then she would play  $l$ , as he would have no other choice than playing  $l$ . This is what we will call coordination failure due to a first-order belief. The second example is that she thinks that his type is high, he thinks she is a high-type player but she thinks that he thinks

<sup>4</sup>Or, in the words of Luce and Raiffa [1957] and Schelling [1960] a *solution in the strict sense*.

<sup>5</sup>In this paper, we follow the notion for coordination failure of Van Huyck, Battalio, and Beil [1990], i.e., the failure to coordinate on the best achievable outcome. That means, even if two high-cognition players coordinate on a Pareto-inferior equilibrium we will call it coordination failure.

her type is low. Again, she would play  $l$  as she thinks that he will play  $l$ . Here we have a coordination problem due to her second-order belief. Therefore, even if both players have the ability to coordinate on the best equilibrium, they might end up failing to coordinate on the better equilibrium  $(h, h)$ .

The existence of infinitely many levels of beliefs and that a “bad” belief at any level makes the player switch to the “bad” strategy  $l$  makes one wonder, if, even with a high fraction of high-cognition players, coordination on the good equilibrium  $(h, h)$  is possible.

Therefore, the first main question this paper addresses is if coordination on the good equilibrium can be expected even in the absence of common knowledge. The second question is if systematic underestimation of other players’ skills can be a source of coordination failure, or if first-order beliefs matter. The third and last question is if higher-order beliefs, e.g. if she thinks that he thinks that she is a low type, are a possible cause for coordination failure or if these levels of reasoning are too complex and play no significant role in coordination games.

The concepts of coordination games and higher-order beliefs will be formalized in the following section and the experimental design will be explained in Section 3.4.

### 3.3 The model

We begin by defining a pure coordination game for two players.

**Definition 1** (Pure coordination game). A *pure coordination game* is a game with 2 players, who each have access to  $m$  different actions  $(\{a_1, a_2, \dots, a_m\})$ . In this game payoffs of a player  $i$  are defined as

$$u_i(a_i, a_j) = \begin{cases} x_i & \forall i, j : i = j \\ 0 & \text{otherwise} \end{cases}$$

with  $x_m > x_{m-1} > \dots > x_1$ .

This means that each player can choose from the same set of actions and whenever they have picked the same action they get the same

payoff and if they don't manage to coordinate their actions, both get nothing. Furthermore, there is a Pareto ordering of these equilibria. Figure 3.2 shows an example of a pure coordination game with three possible actions.

	$a_1$	$a_2$	$a_3$
$a_1$	1, 1	0, 0	0, 0
$a_2$	0, 0	2, 2	0, 0
$a_3$	0, 0	0, 0	4, 4

Figure 3.2: A pure coordination game

Let us now introduce cognitive differences into this pure coordination game. For the sake of simplicity, we are only introducing two cognitive types, a low-cognitive type and a high-cognitive type. The latter has access to a “better” strategy, which is not available to the low type. Furthermore, the low type is unaware of the existence of the high type, as proposed by Bacharach [1993].

This means that the low type has no complete knowledge about the structure of the game and therefore common knowledge about it is, as long as there is at least one low cognition player, not possible.

**Definition 2** (Pure coordination game with cognitive differences). A *pure coordination game with cognitive differences* is a game with 2 players. Each of the players has a type  $t_i \in \{low, high\}$  and has access to different strategies, depending on his type  $t_i$ . The types of a player are her private information. Low cognition players have access to  $\{a_1, a_2, \dots, a_{m-1}\}$  whereas high cognition players also have access to the action  $a_m$ , i.e. to  $\{a_1, a_2, \dots, a_m\}$ . Furthermore, low cognition players have no knowledge about the existence of the high type or action  $a_m$ .

In this game payoffs of a player  $i$  are defined as

$$u_i(a_i, a_j) = \begin{cases} x_i & \forall i, j : i = j \\ 0 & \text{otherwise} \end{cases}$$

with  $x_m > x_{m-1} \geq \dots \geq x_1$ .

These cognitive differences can also be thought of as symmetry constraints on attainable strategies, as proposed by Crawford and Haller [1990] and further developed by Blume [2000] and Alós-Ferrer and Kuzmics [2013]. Here, the high-cognition player has less symmetry constraints and has therefore more attainable strategies.

In the experiment we are using the notion of cognitive differences as proposed by Blume and Gneezy [2010] (a generalization of Bacharach's variable frame theory, using different symmetry constraints on the attainable strategies as used in Blume [2000]).

For a formal description of the belief hierarchy of these games, we would like to refer to Section 3.A. However, we believe for understanding the results of this work, the idea conveyed in this section should suffice.

### 3.4 Experimental design

Measuring higher-order beliefs is very complicated, as there is an "uncertainty principle" (as already discussed by Blume and Gneezy [2010]) at work; i.e., it is hard to measure beliefs without introducing or changing them.<sup>6</sup> Furthermore, introducing absence of common knowledge is difficult; When told that they are given a random number, subjects usually assume that it is drawn from a uniform distribution. Explicitly stating that the distribution is unknown leads to a myriad of other problems. Subjects could, for example, assume a strategic selection of the distribution by the experimenter. Finally, we need to have some sort of control over the fraction of high-cognition players, so that the action only available to the high-cognition players is the one with the highest expected payoff (see Section 3.B).

We solve all three problems by utilizing Blume and Gneezy's [2000] 5-sector disc. This is a disc with 5 equally large sectors on it, 2 black and 3 white, as depicted in Figure 3.3.<sup>7</sup>

---

<sup>6</sup>Either by making the subjects realize that there might be something like a higher-order belief or by them trying to be a good subject (Orne [1962]). A more extensive discussion of this uncertainty principle can be found in Section 3.C.

<sup>7</sup>There is a second version of this disc, with a significantly harder to find distinct

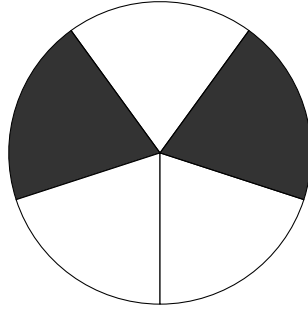


Figure 3.3: 5-sector-disc

The disc has the same sectors on the front- and backside of the disc and can be flipped and rotated.

As the disc can be flipped, the subjects face symmetry constraints and can therefore not distinguish all five sectors. These symmetries cannot be overcome and therefore not all Nash equilibria are possible given the particular frame. Only certain “attainable” equilibria are possible, as defined originally in Crawford and Haller [1990], and further developed by Blume [2000] and Alós-Ferrer and Kuzmics [2013].

The property of this disc which is most important for this paper is that it has a single distinct white sector: The sector adjacent to both black sectors (Figure 3.3).<sup>8</sup>

For the subjects there are then, in principle, three distinguishable sets of sectors: the black sectors (B), the uniquely identifiable white sector (D), and the other white sectors (W’).

The key assumption behind the experiment (and also behind Blume and Gneezy [2000] and Blume and Gneezy [2010] and very much supported by their findings), is that not all subjects realize that there is a uniquely identifiable sector, which leads to two different cognitive types, the high type, who can identify the distinct sector, and the low type, who cannot.

---

sector, with adjacent black sectors. However, for this disc, the fraction of players who were able to identify the distinct sector is too small (i.e., not satisfying the conditions derived in Section 3.B). Some choice data from this “hard disc” can be found in the Online Appendix.

<sup>8</sup>More about the properties of this disc can be found in Blume and Gneezy [2000].

The low type then faces an additional symmetry constraint and has only two distinguishable sets of sectors to choose from: One of the two black sectors (B) or one of the three white sectors (W).

Note that the lower type has no knowledge about the existence of another type or the distinct sector.

The subjects then played three treatments in a random order without feedback after hearing and reading the instructions and completing an extensive quiz:<sup>9</sup>

The **Self Treatment** in which the subject gets the disc twice, every time randomly turned and rotated, and gets £5 if she picks the same sector twice.

In the **Prediction Treatment** one subject (she) is told that another subject (he) plays the *Self Treatment* (with a possibly differently turned and rotated disc). She has to pick one sector and every time he picks the sector she picked, she gets £2.5.<sup>10</sup>

Finally, the **Coordination Treatment**, in which two players pick simultaneously a sector on a (randomly turned and rotated) disc and, if both players pick the same sector, both receive £5.

### 3.4.1 Predictions

How can we use this design to test the three initial questions stated in the introduction? Let us have a look how we expect low- and high-cognition players to behave in the three different treatments.

In the Self Treatment a high-cognition player has 9 possible choices: She can pick any of three actions ( $D$ ,  $B$ ,  $W'$ ) in the first stage and then pick any of the three actions in the second stage. This decision problem for the high-cognition player has a unique optimal solution: pick the distinct sector twice, giving her a probability to win of 1.

---

<sup>9</sup>For the complete instructions and a description of the quiz see the Online Appendix.

<sup>10</sup>Adding another treatment in which subjects have to predict what another subject does in the Prediction Treatment would, in theory, allow to explicitly check for second-order beliefs (or, when repeating this any higher-order belief). However, don't believe this will work with the 5-sector disc, as it probably requires too much attention and mental effort which most subjects might not be willing to exert.

A low-cognition player is only aware of four possible choices: He can pick  $B$  or  $W$  in the first stage and then pick  $B$  or  $W$  in the second stage. The low-cognition player also has a unique optimal choice: pick  $B$  in both stages, giving him a probability to win of  $\frac{1}{2}$ .

Therefore, we would expect a high-cognition player to choose the distinct sector twice and a low-cognition player to pick a black sector twice.

In the Prediction Treatment, the action taken by a subject should only depend on her type and her first-order belief about the type of the other player. A low-cognition player will always choose  $B$ , whereas a risk-neutral, high-cognition player should pick  $D$  if his belief that the other player is also of the high type is at least  $\frac{1}{3}$  and  $B$  otherwise.<sup>11</sup>

The coordination treatment is best depicted as a bi-matrix game with three (for the high-cognition player) and two (for the low-cognition player) pure strategies, with winning probabilities as depicted in Figure 3.4 and Figure 3.5. We expect a low-cognition player to play  $B$ , as it is the payoff- and risk-dominant equilibrium, whereas a high-cognition player's choice depends on her belief hierarchy: If anywhere in her complete hierarchy a belief lower than  $\frac{1}{3}$  (or  $\frac{1}{2}$  for very risk averse players) that the other player is a high-cognition player or that the other player thinks that she is a high-cognition player, ... (or, in short, that there is no common-p belief among the high-cognition players of  $\frac{1}{3}$  or higher, that both players are high-cognition players), she will choose  $B$ , otherwise she will choose  $D$ .

	$W'$	$B$	$D$
$W'$	$\frac{1}{2}$	0	0
$B$	0	$\frac{1}{2}$	0
$D$	0	0	1

Figure 3.4: High-cognition player winning probabilities

	$W$	$B$
$W$	$\frac{1}{3}$	0
$B$	0	$\frac{1}{2}$

Figure 3.5: Low-cognition player winning probabilities

Unfortunately, neither the theoretical nor the experimental litera-

<sup>11</sup>Allowing for risk-averse players, this fraction has to be between  $\frac{1}{3}$  and  $\frac{1}{2}$ , depending on the degree of risk aversion.

ture on higher-order beliefs can tell us which of the two will be chosen. Even small variations in the assumptions of theoretical models of higher-order beliefs can generate both equilibria. Therefore, this question seems to be an empirical one, which we are trying to answer in this paper. However, a more detailed explanation of how different models of higher-order beliefs work in our game can be found in Section 3.B.

### 3.4.2 Hypotheses

Using our design, we can formulate three hypotheses to test the three research questions stated earlier. In the following we will use a shorthand for players' strategies such as: "W'W' B D" means that a player selected one of the two white sectors twice in the Self Treatment, one of the black sectors in the Prediction Treatment and the distinct sector in the Coordination Treatment.

The answer to our first question "Is coordination possible?" or, in the words of our model "Do high-cognition players use the first-best strategy  $a_m$  despite the absence of common knowledge?" is suggested by the literature on focal points (e.g., Sugden [1995] or Crawford, Gneezy, and Rottenstreich [2008]) and supported by the experimental literature on coordination games (e.g., Van Huyck, Battalio, and Beil [1990] or Cooper, DeJong, Forsythe, and Ross [1990]):

**Hypothesis 1** (Coordination is possible). *High-cognition players choose in the Coordination Treatment D more often than any other choice.*

We are using a within-subject design to test the hypotheses: Only high-cognition players can identify the best equilibrium, so we don't have to consider other types. We can identify these players with the help of the the Self Treatment. If high-cognition players, i.e., the ones who have been able to identify "D" in the Self Treatment, coordinate on D in the Coordination Treatment we know that coordination is possible, even in the absence of common knowledge.

The next two hypotheses extend on Blume and Gneezy's [2010] hypothesis that "beliefs matter": Hypothesis 2 formalizes the question



“Does coordination fail because some high-cognition players underestimate the fraction of high-cognition players? “ or “Is there coordination failure due to first-order beliefs? “

**Hypothesis 2** (First-order beliefs matter). *There are high-cognition subjects who choose a black sector in the Prediction Treatment and Cooperation Treatment, i.e., play “DD B B”.*

We already know that we can identify players’ types with the help of the Self Treatment. Furthermore, the Prediction Treatment identifies players who think that more than  $\frac{1}{3}$  of the other players can not identify the distinct sector.

Most of the problems in models of higher-order beliefs stem from the fact that there are infinitely many levels of beliefs. However, evidence from the laboratory indicates that people are not able to use higher-order rationality,<sup>12</sup> a requirement for coordination problems due to higher-order beliefs. Furthermore, even in studies of level-k reasoning, where players are framed and incentivized on using higher-order beliefs, players still rarely use high levels of reasoning.<sup>13</sup>

Therefore, the third question, if there is coordination failure due to higher-order beliefs, or if high-cognition players use the first-best strategy  $a_m$  despite the absence of common knowledge, arises naturally:

**Hypothesis 3** (Higher-order beliefs matter). *There are high-cognition subjects who play the distinct sector in the Prediction Treatment and a black sector in the Cooperation Treatment, i.e., play “DD D B”.*

Our design allows for another robustness check: There is an attainable strategy which is very similar to the one we use to identify first- and higher-order beliefs: “DD B D”. This strategy will only be chosen if players believe that their partner is of the low type, but still plays “D”

---

<sup>12</sup>Kneeland [2015] shows that only about 22% of all players use more than third-order rationality.

<sup>13</sup>In Arad and Rubinstein’s [2012] 11-20 game, 80 % of the players only use 3rd-order beliefs or lower despite the game being designed to facilitate higher-order reasoning.

in the in the Coordination Treatment. This strategy can therefore not be explained using our model.

**Hypothesis** (Robustness check). *“DD B D” is played less often than “DD B B” and “DD D B”.*

### 3.5 Results

The experiment was conducted at the DR@W Laboratory at the University of Warwick using the experimental software "z-Tree" developed by Fischbacher [2007]. 130 subjects were recruited and received payments between £3 and £18. Before showing the results, let us briefly discuss the preliminaries of the experiment design.

The first preliminary is the focality of the distinct and the two black sectors. From the choice data in Figure 3.6 we can see that more than 95% of all players have chosen one of these sectors in the Coordination Treatment. Therefore, the black and distinct sectors seem to be focal in our game. The second preliminary is that there are enough high-cognition players, so that playing the high-cognition exclusive action is a payoff-dominant equilibrium for the players. In Figure 3.7 we can see that 58% of all players have chosen the distinct sector and are therefore considered high-cognition players. Therefore, playing the distinct sector would maximize the expected utility of high-cognition players in a game where the type distribution is common knowledge among high types independently of the degree of risk aversion (see Section 3.B). We can also see that the second most frequently observed behavior is choosing a black sector twice, whereas choosing a white sector twice (which includes choosing the distinct white sector once and another white sector once) and picking one black and one white sector (labeled “Other”) was very rare.

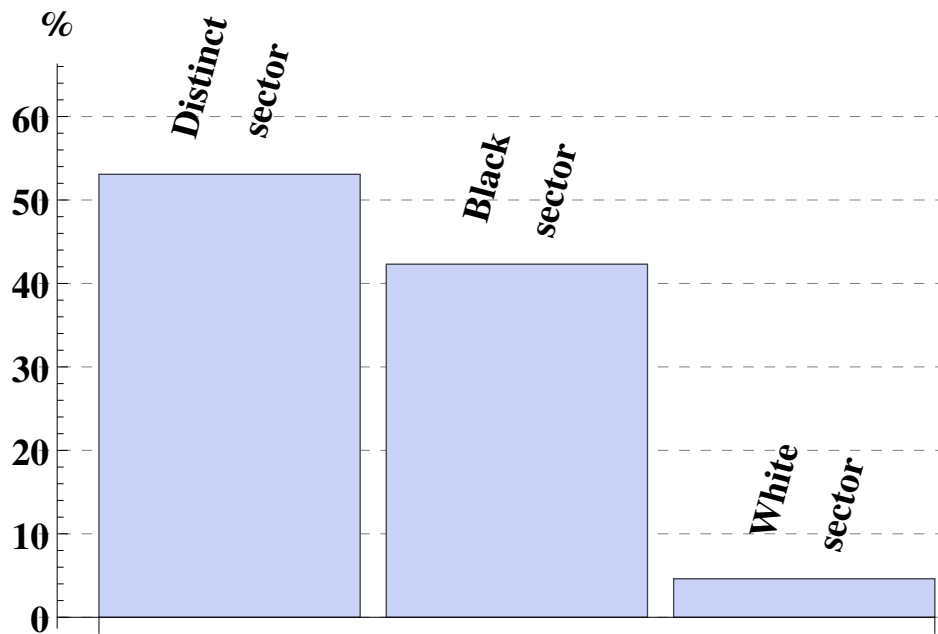


Figure 3.6: Results of the Coordination Treatment

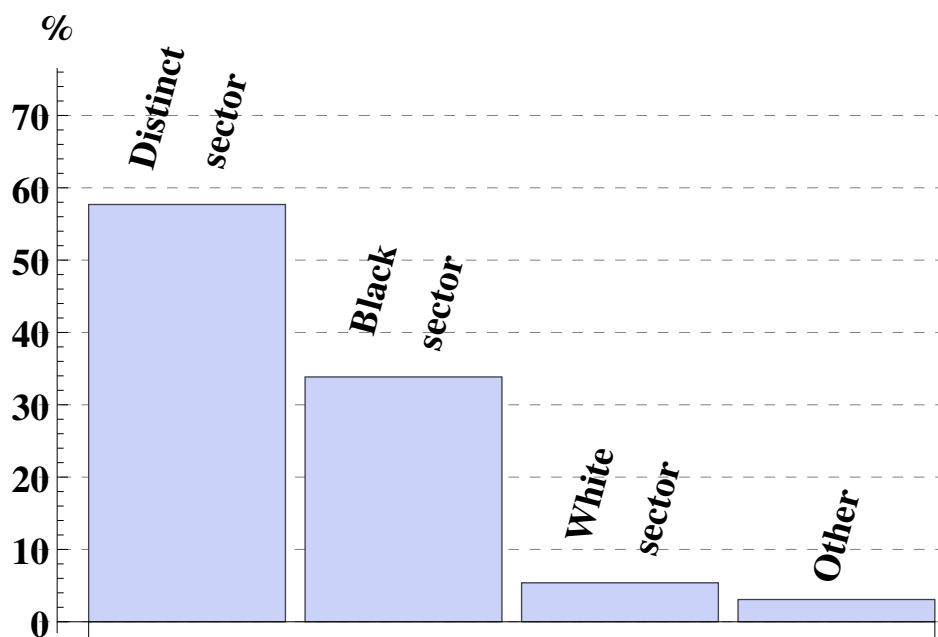


Figure 3.7: Results of the Self Treatment

These results are in line with Blume and Gneezy's [2010] results where around 52% (58% in our experiment) have been able to identify

the distinct sector and around 23% (34%) have chosen the black sector. We contribute the significantly lower level of noise (8% vs 25%) to the extensive instructions and the quiz we conducted before the experiment.<sup>14</sup>

Due to the lower level of noise we are, unlike Blume and Gneezy [2010], able to use a within-subject design, in which each player has access to 625 possible strategies.<sup>15</sup> Of these strategies we consider 96.32% as “noise”.<sup>16</sup> As the number of strategies which support our hypothesis are very low (1, 4 and 2 out of 625), the probability that someone chooses them by mistake is very low. For a detailed overview of all possible strategies and how we categorize them see Table 3.1.

Description	Hypothesis	# of strategies	Proportion
DD D D	1: Coordination is possible	1	0.16%
DD B B	2: First-order beliefs matter	4	0.64%
DD D B	3: Higher-order beliefs matter	2	0.32%
BB B B	(Low-cognition players)	16	2.56%
“Noise”	-	602	96.32%
WW W W	(part of “Noise”)	80	12.80%

Table 3.1: Overview of the strategies

Given the preliminaries, we can test hypotheses 1 through 3.

**Hypothesis 1** (Coordination is possible). *High-cognition players choose in the coordination treatment D more often than any other choice.*

The choice data from our experiment confirms this hypothesis. In Figure 3.8 we can see that 80% have chosen the strategy “DD D D”. As this strategy represents only 0.16% of all available strategies (or 4% when excluding the Self Treatment), we can reject the null hypothesis

<sup>14</sup>For the instructions and an overview of the quiz see the Online Appendix.

<sup>15</sup>We are here ignoring the order in which treatments are played.

<sup>16</sup>This noise includes not only players not understanding the experiment or behaving randomly but also “Eureka”-learning (which was a big problem in Blume and Gneezy [2010], see Section 3.C), making a mistake (e.g., picking a not distinct white sector instead of the distinct sector, a mistake, which both of the authors made multiple times while testing the experiment) and beliefs of low-cognition players.

of this high level of coordination being a result of random play ( $p < 0.00001$ ).

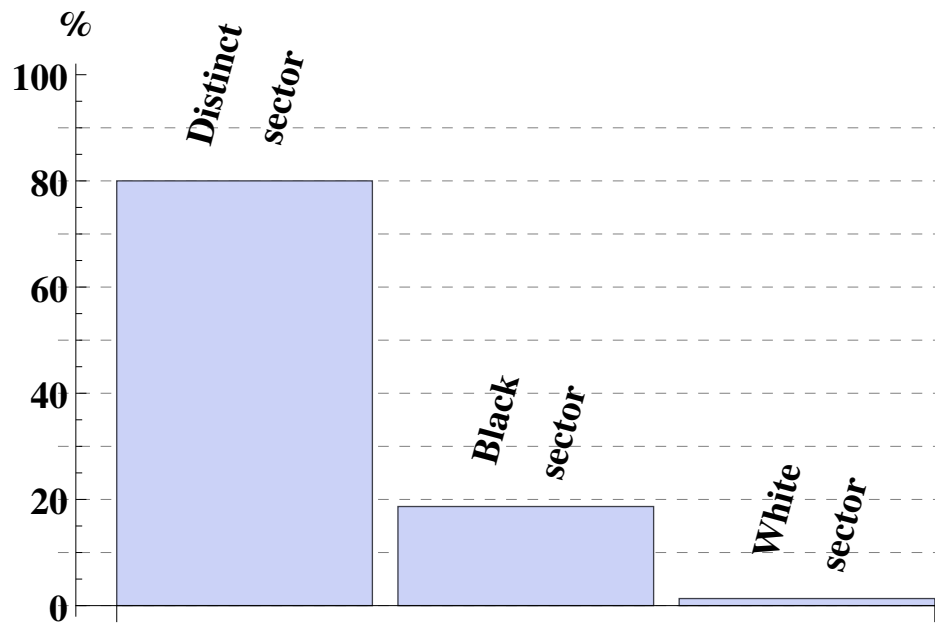


Figure 3.8: Results of the Coordination Treatment (high-cognition players)

Blume and Gneezy [2010] claim that “beliefs matter” and we test in Hypothesis 2 if there are subjects whose pessimistic beliefs about the other players’ skills lead to coordination failure.

**Hypothesis 2** (First-order beliefs matter). *There are high-cognition subjects who choose a black sector in the Prediction Treatment and Cooperation Treatment, i.e., play “DD B B”.*

Our data confirms this hypothesis. Figure 3.9 shows us the results of all players, Figure 3.10 of the high-cognition players. In these figures we can see that about 9% of the high-cognition players (or 5% of all players) have a first-order belief problem, leading to coordination failure. As the fraction of strategies leading to this conclusion is very small (0.64%) we can reject the null hypothesis that this result is due to chance ( $p < 0.00001$ ).

But do players really use higher-order beliefs in this type of games? Hypothesis 3 tests for this question.

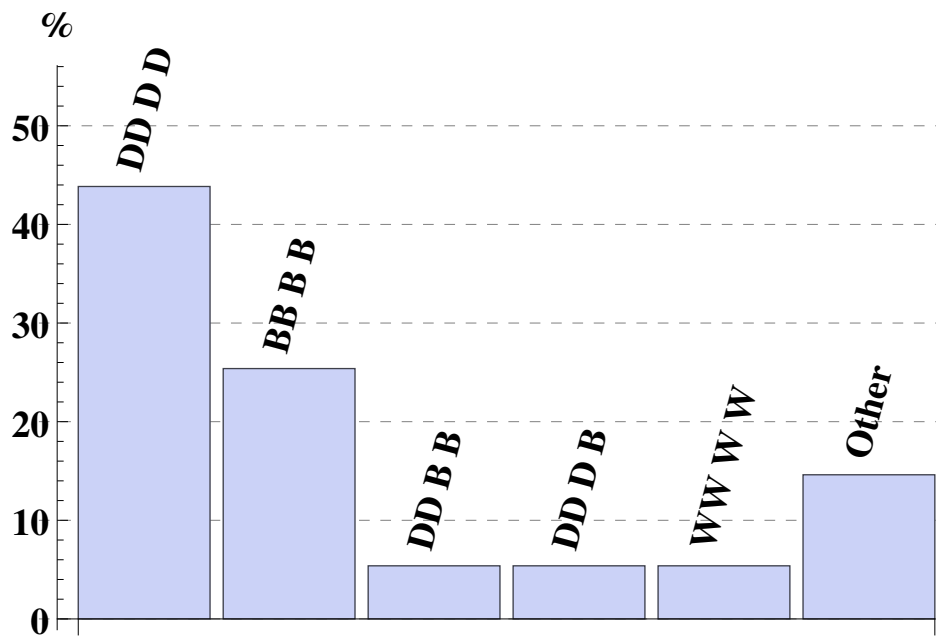


Figure 3.9: Used strategies

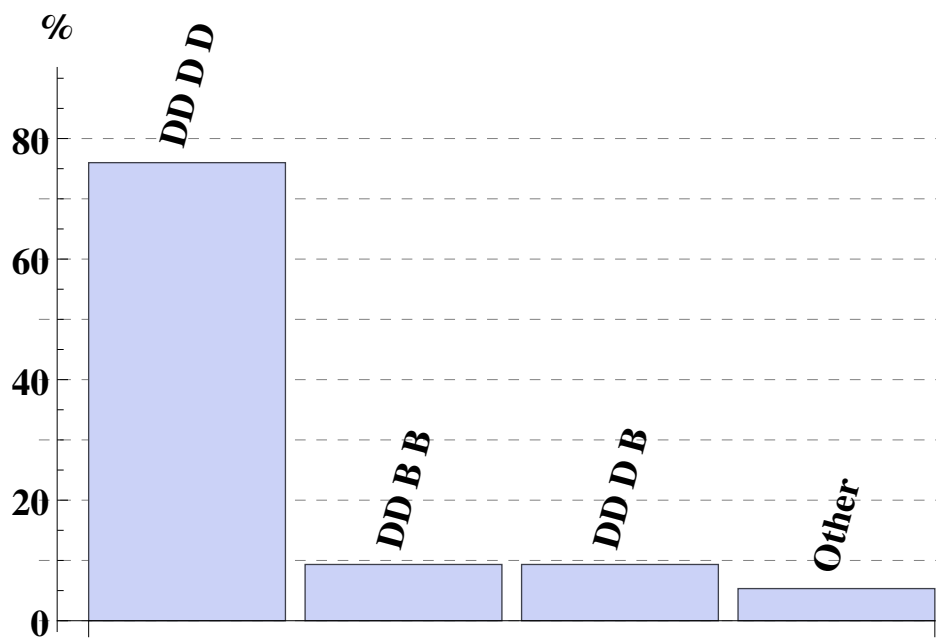


Figure 3.10: Used strategies (high-cognition players)

**Hypothesis 3** (Higher-order beliefs matter). *There are high-cognition subjects who play the distinct sector in the Prediction Treatment and a black sector in the Cooperation Treatment, i.e., play “DD D B”.*

From Figure 3.9 and Figure 3.10 we can see that there are high-cognition players who think that their partner is with a high probability of the high type, they, however, still think there are coordination problems. Again, we can reject the null hypothesis at the 1% level ( $p < 0.00001$ ).

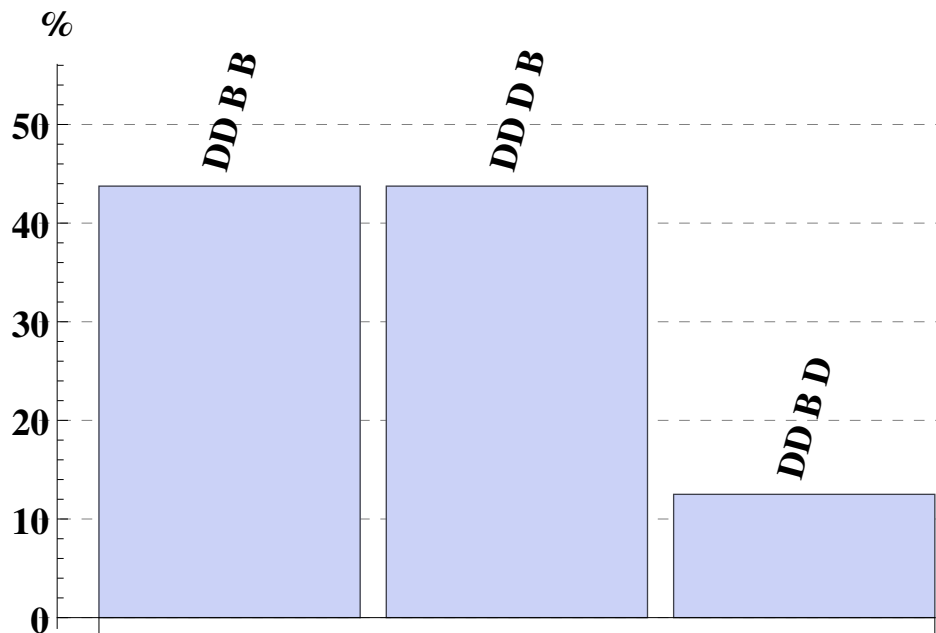


Figure 3.11: Robustness check

**Hypothesis** (Robustness check). *“DD B D” is played less often than “DD B B” and “DD D B”.*

All these results are statistically significant at the 1% level, however, our design allows for another robustness check: There is a strategy which should not be played by rational players: “DD B D”, which is about as likely to be picked at random as “DD B B” and “DD D B” but can not be explained by our model. Figure 3.11 shows us that only 2 subjects have chosen this strategy.

We expected to have significant order effects, as in Blume and Gneezy [2010]. However, it turns out, that the only robust order effect is a weak effect in the Self Treatment (i.e., more subjects have been able to choose the distinct sector twice later in the experiment).<sup>17</sup> We attribute this to a small change in design. We have explained every treatment before the experiment started and we have conducted a quiz (Section 3.E.2), testing if the instructions have been understood. This probably lead to “Eureka learning” before instead of during the experiment.

### **3.6 The role of beliefs in the fight against female genital mutilation**

More important applications than the search for ones husband or wife in a department store are suggested by recent studies in sociology and development studies, for example by Bicchieri [2005].<sup>18</sup> She claims that common knowledge plays a significant role in the fight against female genital mutilation (FGM).<sup>19</sup> FGM is practiced in, predominately African, communities and is required in many of these communities to find a husband and to prevent social exclusion. Despite being very dangerous and unnecessary, it has a long standing tradition and is, in areas where it is still practiced, very common. It is estimated to effect up to 200 million women in 2016 (UNICEF [2016]). In game theoretic terms the problem is one of equilibrium selection: There is one equilibrium in which everyone accepts and uses FGM and one in which no one does. The latter equilibrium is, given enough knowledge about the subject, clearly better for everyone, but we still observe the former equilibrium in many communities.

How is that problem related to higher-order beliefs? This game might be modeled as a coordination game without common knowl-

---

<sup>17</sup>For the full analysis of order effects see Section 3.C.

<sup>18</sup>More examples in which common knowledge plays an important role can be found in Chwe [2013].

<sup>19</sup>Most studies, however, don’t use the terms “beliefs” or “common knowledge” but describe this concepts in their own words, frequently restricting their attention to first-order and therefore ignoring higher-order beliefs.



edge about the payoffs. While, from our point of view, the no-FGM equilibrium is clearly better, there are societies who disagree and belief in the (perceived) benefits while ignoring the risk. Using this model, the game is very similar to the game we played in our experiment:<sup>20</sup> The high type knows the real payoffs (i.e., that the non-FGM equilibrium is better than the FGM equilibrium), whereas the low type thinks that the opposite is true. An important tool in the fight against FGM is to inform people about the dangers and (lack of) benefits of it. The equivalent in this model would be to chance the type of these families from low to high. However, from our experiment we know, that just changing the type might not be sufficient, as some people have beliefs that still lead to coordination failure. Here we can see that differentiating between first-order beliefs and higher-order beliefs matters. While taking care of first-order beliefs can be easily done by explaining, in addition to the risks and lack of benefits of FGM, that other families have also been (or will be) educated on this matter and that most families agree with these results, higher-order beliefs are not as simple. To prevent coordination failure due to higher-order beliefs everyone has to be gathered and the education has to be done in very large groups, which is significantly more difficult and expensive.

But how important is this issue really, considering that the model is very simplistic (we are, e.g., ignoring all dynamic issues) and that generalizing from a lab experiment, conducted with university students in Europe, to rural areas in Africa requires much more research and data from (field) experiments. Studies like Bicchieri [2005] suggest that higher-order beliefs might play a role; She claims that common knowledge of this education plays an important role because negative beliefs about the opinion of the other members of a community might prevent a coordination on the better equilibrium (i.e., the one without FGM): Even if I am convinced that this practice should be abolished, I might still partake in it, to prevent my daughters from being excluded from the community, as the others might not be convinced (i.e., my

---

<sup>20</sup>The absence of common knowledge here is, however, over the payoffs, not the available actions.

first-order belief is that others have not been educated). I also might think that others will continue this practice because they think I wasn't educated (i.e., because of my second-order belief) and so on.

That means, that just educating a family (or, in game theoretic terms: changing their type) does not necessarily lead them to change their stance on FGM. But is there any evidence that families use beliefs? Mackie [1996] and Mackie and LeJeune [2009] have compared the old Chinese tradition of foot binding<sup>21</sup> and FGM and pointed out that both are similar: Both are required to find a husband, while being very painful and dangerous without having any known benefits. Furthermore, they have a long-standing tradition (both can be traced back more than 1000 years) and were widely spread in their respective cultures. However, around 1910, foot binding has dropped in certain parts of China from 99% to under 1% prevalence over the course of just 20 to 30 years, without any change in policy (Gamble [1943], Keck and Sikkink [1998]), whereas even a combined effort of the UN, several NGOs and governments over the last 40 years resulted only in a moderate decline from about 51% to 37% of women effected by FGM in certain countries (UNICEF [2016]). Mackie [1996] claims that the main difference is the method of information transmission: In China, societies have been founded in which members publicly pledged to not bind their daughters' feet and to prevent their sons from marrying women with bound feet, whereas the effort to prevent FGM was mainly focused on changing the laws and educating the people about the dangers and problems. The societies fighting foot binding made the education and position of the families common knowledge whereas most organizations fighting FGM focused on changing the opinion of the families without changing the higher-order beliefs much.

But also between projects fighting FGM there have been differences. Tostan, a Senegal-based NGO, has, according to World Bank Group [2012] successfully reduced the number of FGM in some parts

---

<sup>21</sup>Foot binding was a Chinese practice of bending and binding the four lower toes of young girls tightly under the foot.

of Senegal significantly. So, why did Tostan succeed where others have failed? They claim that not only education but " [...] public declarations are critical in the process for total abandonment [of female genital cutting.] " (Tostan [2016]) and are supported by World Bank Group [2012] who emphasizes that education together with public discussion and public declaration was an important factor in Tostan's success.

These examples suggest that beliefs might play an important role, as the more successful campaigns against foot binding and FGM also addressed higher-order beliefs by introducing common knowledge whereas others who focused on pure education have been less successful. However, it is not clear that common knowledge is required to achieve coordination. It might be sufficient to explain that others have also been educated (i.e., to take care of the first-order beliefs), which would be much cheaper than providing common knowledge.

However, from our results we know that ignoring the higher-order beliefs can have severe negative consequences. Our results can explain why education without considering problems due to higher-order beliefs can have some effect but they can also explain why NGOs like Tostan have significantly more success. Furthermore, these results give reason to belief that only explaining if others have been educated and are against FGM (i.e., changing the first-order beliefs) might not be sufficient and making this education common knowledge might be necessary to achieve all possible benefits from it.

However, the results from this experiment conducted with students at a European university should of course not be generalized to explain behavior in small rural communities without further research but gives us reason to belief that higher-order beliefs do matter.

### 3.7 Conclusion

We have seen that, in this game, absence of common knowledge was not enough to prevent subjects to choose the Pareto-dominant equilibrium strategy, as 76% of the high-cognition players have done so. However, we still have a fraction of players who have beliefs that lead

to coordination failure (around 18%) and of these only half could be attributed to first-order beliefs.

Of the models of higher-order beliefs discussed in Section 3.4.2 and Section 3.B, only "assuming common knowledge"<sup>22</sup> or a common p-belief among high-cognition players about the type distribution were able to explain coordination on the payoff-dominant equilibrium. However, these assumptions can not explain any coordination failure due to beliefs, as the beliefs are fixed by the model, whereas the models which can explain this type of coordination failure predict playing the payoff-dominated strategy.

Therefore, as we have observed a coordination rate of about 76%, assuming common knowledge (or a common p-belief among high-cognition players) might be the best tractable approximation available in coordination games without common knowledge, depending on the focus of the research.

Our work opens up some questions for future research: Can these results be generalized to other populations and environments? Are there certain parts of the populations who are more likely to exhibit first- or higher-order beliefs which lead to coordination failure? Are there other, maybe easier methods to make something common knowledge? Furthermore, it might be worthwhile to check more general structures of higher-order beliefs or if non-equilibrium models like Strzalecki [2014] or Kneeland [2016] can explain this phenomenon better.

---

<sup>22</sup>Meaning that one assumes that high-cognition players have common knowledge about the type distribution among themselves.

## Appendix 3.A Belief hierarchies

Let  $B_i^0 := T_j$  and  $B_i^k = T_j \times \Delta(B_i^{k-1})$  with  $\Delta(B)$  being the space of probability measures on  $B$  and  $\Delta(X)$  being the space of probability measures on the Borel field of  $X$ , endowed with the weak topology. Using this notation, we can define a belief hierarchy as follows.

**Definition 3** (Belief hierarchy). A  $k$ -th order belief is defined as

$$b_i^k \in \Delta(B_i^k)$$

with  $B_i^0 = T_j$  and  $B_j^k = T_j \times \Delta(B_j^{k-1})$

Furthermore, let us set  $b_i^0 := t_i$ .

A belief hierarchy of a player  $i$  is then  $b = \{b_i^0, b_i^1, \dots\}$

We therefore have a first order belief  $b_i^1 \in \Delta(\{low, high\}) = [0, 1]$  and higher-order beliefs  $b_i^k \in [0, 1]^k$ .

Furthermore, we assume these beliefs to be coherent, i.e. that beliefs of different orders do not contradict one another,<sup>23</sup> and that a low-cognition type does not know about higher cognitive types, i.e.,  $b_i^k = 0 \Rightarrow b_i^{k+1} = 0 \quad \forall k \geq 0$ .

This excludes, on the one hand, that a low-cognition player thinks that the other player is a high-cognition player and, on the other hand, that a player has a first-order belief that the other player is of a the high type and a higher-order belief that the player is of the low type.

## Appendix 3.B Equilibrium selection and models of higher-order beliefs

In this section we are going to discuss how different models of beliefs and frequently used assumptions on the structure of higher-order beliefs influence the specific game we analyze.

Using the results from the literature on focal points in coordination games (as discussed in Section 3.1.1), we know that we can restrict

---

<sup>23</sup>I.e., higher-order beliefs of a player mapped onto the space of beliefs of a lower order are the same.

our attention on the two actions with the highest payoffs  $a_{m-1}$  and  $a_m$ . This simplifies the game to a Bayesian game with two types, a low type whose only attainable action is  $a_{m-1}$  and a high type, who has access to  $a_{m-1}$  and  $a_m$ , without common knowledge about the type distribution. Then, we can denote, with a small abuse of notation, the strategy of a player as the action she chooses if she is of the high-type, i.e.,  $a_m$  or  $a_{m-1}$ , knowing that she will play  $a_{m-1}$  if she is of the low type.

Let us first start with the most common assumption, that the *distribution of types is common knowledge*. Then the expected utility of a (risk neutral) high-cognition player is as depicted in Table 3.2, given her and her partners strategies.<sup>24</sup>  $p$  denotes here the probability of a player being of the high type. We can see that the prediction of the model then depends on  $p$ . If the probability of a player being of the high type  $p$  is too low ( $p < \frac{x_{m-1}}{x_{m-1}+x_m}$ ), only  $(a_{m-1}, a_{m-1})$  will be an equilibrium. In this paper we are going to assume that  $p \geq \frac{x_{m-1}}{x_{m-1}+x_m}$  which makes sure that the “better” equilibrium always exists.<sup>25</sup> For risk-averse players, it is required that  $p \geq \frac{u(x_{m-1})}{u(x_{m-1})+u(x_m)}$ , so we know that as long as  $p \geq \frac{1}{2}$  the high-type equilibrium always exists, independently of the degree of risk aversion. Furthermore, if the equilibrium exists, it is payoff dominant.

	$a_{m-1}$	$a_m$
$a_{m-1}$	$x_{m-1}, x_{m-1}$	$(1-p)x_{m-1}, 0$
$a_m$	$0, (1-p)x_{m-1}$	$px_m, px_m$

Table 3.2: Expected utilities of two high-cognition players

Therefore, the prediction of assuming that the *distribution of types is common knowledge* is that, for a high-enough  $p$ , we should expect full cooperation.

Monderer and Samet’s [1989] *common p-belief* is a generalization of the concept of common knowledge and generates, in this model, the

<sup>24</sup>In the analysis we restrict our attention to risk-neutral players. However, the analysis for the case of risk-averse players is analogous and the experimental results are valid for every possible degree of risk aversion.

<sup>25</sup>In the experiment this assumption requires  $p > \frac{1}{3}$ . As the fraction of high-cognition players is 58%, this assumption is not problematic.

same predictions as assuming that the distribution of types is common knowledge, given a high-enough  $p$ .

The game we are analyzing is very close to the original description of a global game as introduced by Carlsson and Van Damme [1993]. Written down as in Table 3.2 it is a very similar game as the main example used in Carlsson and Van Damme [1993]. Therefore, we know that, given  $\frac{x_{m-1}}{x_m} \leq p \leq \frac{2x_{m-1}}{x_m+x_{m-1}}$  (i.e.,  $(a_m, a_m)$  is still a Nash equilibrium but  $(a_{m-1}, a_{m-1})$  is risk dominant),  $(a_{m-1}, a_{m-1})$  will be the only rationalizable solution to the global game. Furthermore, Hellwig [2002] shows that higher-order uncertainty about preferences leads to results similar to Carlsson and Van Damme's [1993] higher-order uncertainty about payoffs, i.e., coordination on the "less risky" equilibrium.

Rubinstein [1989] shows that truncating common knowledge at any finite level is equivalent to the situation without any common knowledge at all and therefore suggests that players choose the same strategy  $a_{m-1}$ .

Weinstein and Yildiz [2007a] establish a condition, called "global stability under uncertainty" which implies that the change in equilibrium actions is small in the change of  $k$ th-order beliefs and higher. Therefore, under this condition, equilibria can be approximated by the equilibrium with lower-order beliefs. Unfortunately, pure coordination games do, in general, not fulfill the conditions for "global stability under uncertainty" as the best responses are very sensitive to every order of beliefs and even a small change in some higher-order belief might make a player change from  $a_m$  to  $a_{m-1}$ .

## Appendix 3.C Order effects

Earlier, we have briefly discussed an uncertainty principle, in which higher-order beliefs can not be measured without inducing them. This theory is related to the "good subjects hypothesis" (Orne [1962]) according to which some subjects try to figure out the research question and then change their behavior to confirm said hypothesis. However, in this case the difference is more subtle: As soon as they realize that

Model	Coordination	First-order belief coordination problems	Higher-order belief coordination problems
Common knowledge	Full coordination	No	No
Common p-belief	Full coordination	No	No
Global games	No coordination	Yes	Yes
Almost common knowledge	No coordination	No	Yes.

Table 3.3: Models of higher-order beliefs

there is a higher-order belief problem, they might overestimate it.

Blume and Gneezy [2010] have encountered a different case of this uncertainty hypothesis. "Having a player play against himself may trigger an insight that switches a player from low to high cognition ("Eureka!" learning). There may be an uncertainly principle at work here in that we cannot measure a player's cognition without altering it." (Blume and Gneezy [2010]) This suggests, that the order of treatments might be important. Therefore, we implemented a random order. However, it turns out that we have (almost) no order effect, as can be seen in Table 3.4. The only statistically significant effect is that, if the self treatment was the first treatment, there was a significantly higher number of "Other" results than when it was the second ( $p = 0.0062$ ) or third treatment ( $p = 0.0139$ ). Furthermore, the distinct sector was played more often in the coordination treatment if it was the second than the first treatment ( $p = 0.0277$ ), however, there were no significant effects when comparing the first and third and the second and third.<sup>26</sup> The former has a intuitive explanation (i.e., practicing the task makes it less likely to make a mistake) whereas the later is considered to be a type II error by the authors.

The question now is, why did Blume and Gneezy [2010] encounter strong "Eureka!"-learning effects whereas we had (almost) no signif-

---

<sup>26</sup>Using the one-tailed Fisher's exact test.



<b>Treatment</b>	<b>Self</b>			<b>Prediction</b>			<b>Coordination</b>		
<b>Order</b>	<b>DD</b>	<b>BB</b>	<b>Other</b>	<b>D</b>	<b>B</b>	<b>W</b>	<b>D</b>	<b>B</b>	<b>W</b>
<b>1st</b>	18	10	8	23	15	3	22	22	3
<b>2nd</b>	32	24	2	20	16	2	24	10	0
<b>3rd</b>	25	10	1	32	15	4	23	15	5

Table 3.4: Order effects of the different treatments

ificant effect. The authors attribute this to the fact that we used more extensive instructions and a quiz to make sure the instructions were understood. More importantly, the participants were instructed in all three treatments before they played the first game which most likely triggered the learning before the first decision, whereas in Blume and Gneezy [2010] the instructions for the second treatment were distributed after completion of the first treatment.

## Appendix 3.D Data

For the complete data set please refer to the Online Appendix.

[illegible]

Table 3.5: Strategic choice data

## Appendix 3.E Online Appendix

### 3.E.1 Instructions

Welcome to this experiment in economic decision making. It will take approximately 60 minutes. First of all, please check that the number on the card handed to you matches the number on the cubicle that you are seated in and that your mobile phones are turned off.

Before we start, we will explain the rules of this experiment. You will also find these rules on the paper provided, so you can read along and check again during the experiment. If you have any questions, please do not speak up but raise your hand and we will come to you and answer your question privately.

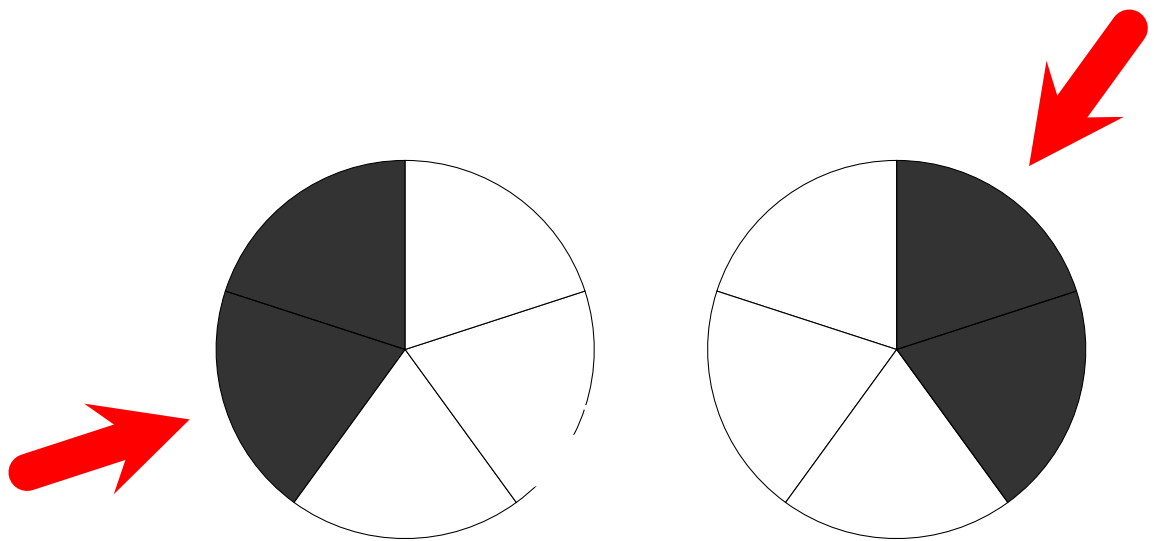
From now on, please do not talk, and listen carefully. In this experiment you will earn a minimum of £3, and potentially up to £18. How much money you earn will depend on your decisions and those of the other participants. Your reward will be paid out at the end of the experiment. None of the other participants will know how much money you made.

In this experiment you will be asked to make decisions related to a disc that has 5 sectors, similar to the disc provided to you. The disc has two identical sides. Your goal will be to pick the same sector twice (more on that later). During this experiment the disc will be flipped and/or rotated randomly.

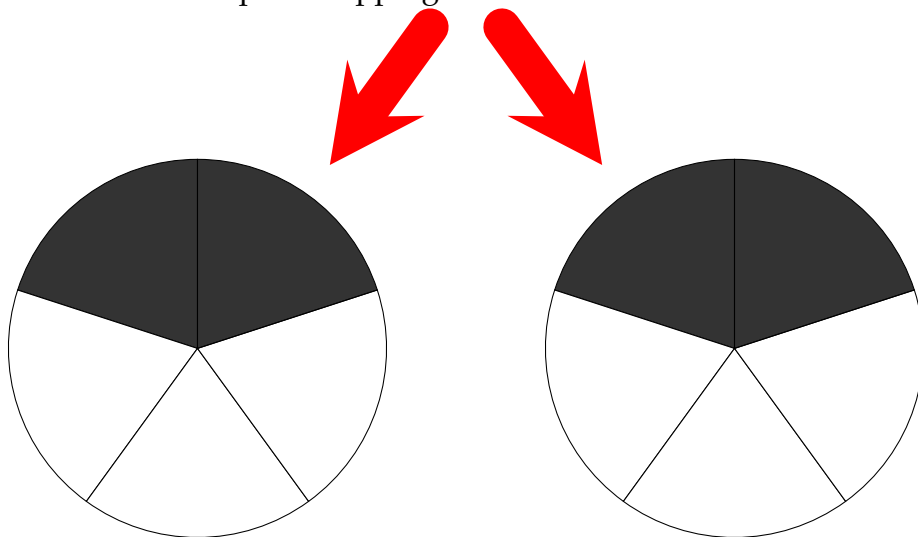
Pictures on page 2 illustrate rotation and flipping. Since you will not be told if the disc was flipped and/or rotated, it might even be the case that disc looks exactly the same though sectors have changed their positions.

The arrow tracks one specific sector that changes its position as the disc is rotated and/or flipped.

This is an example of rotating the disc by two sectors:



This is an example of flipping the disc:



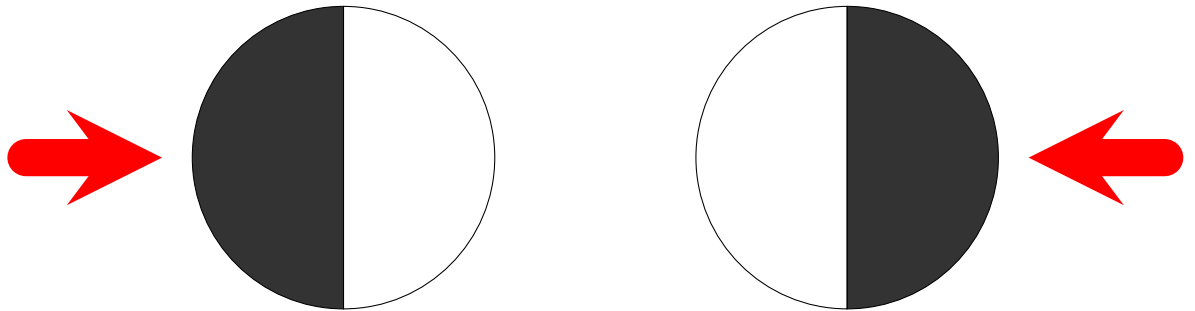
In the experiment the disc will be surrounded by the letters A, B, C, D, and E. **These labels are not part of the disc!** They are only included to allow you to choose a sector.

In the experiment you will make decisions in the following environments (the order will be chosen randomly):

**(Self Game)** You will be asked to pick a sector twice; first you choose a sector; then the disc might be flipped and/or rotated. After this you are shown the same disc and have to choose a sector again. You will not observe the flipping/rotation of the disc. If you manage to guess the same sector twice, your payoff will be £5. Otherwise, you

will receive 0. Therefore, to earn more money you want to maximise your chances to pick the same sector twice.

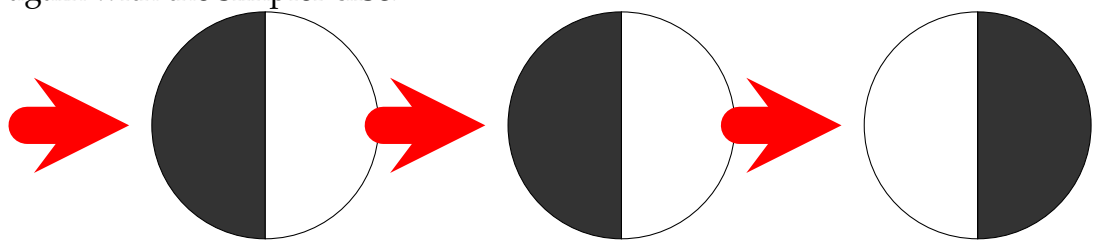
Here is an example of the choices made in a **Self Game**, using a simpler disc with only 2 instead of 5 sectors:



First you picked the black sector; then you picked the black sector again. Therefore, you pick the same sector twice and earn £5.

**(Prediction Game)** You are matched randomly with another person and you have to guess the choice of this person, while she plays the **Self Game**. First, you choose a sector on the disc; each time the other person picks the sector you chose, you will receive £2.5. As the other player picks twice in the **Self Game**, you can earn £0, £2.5 or £5 in this situation, depending on your and the other person's choice. Therefore, to earn more money you want to guess what the other player is playing in the **Self Game** described above.

Here is an example of the choices made in a **Prediction Game**, again with the simpler disc:

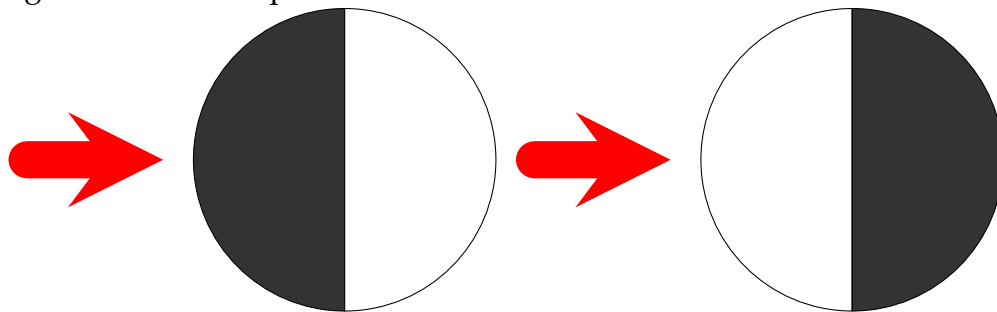


First you picked the black sector. The other player then plays the **Self Game**. He first picks the black sector and therefore you earn £2.5. Then he picks the white sector and therefore you earn £0. Thus you earn £2.5 in total.

**(Coordination Game)** You are matched randomly with another person and both of you are asked to pick a sector on the disc simultaneously. Both of you know that you play the **Coordination Game**. You

both see the same disc but possibly differently flipped and rotated. If both of you pick the same sector, then your payoff will be £5. Otherwise, you will receive £0. Therefore, to earn more money you want to guess the sector your partner is picking here, while he is trying to do the same.

Here is an example of the choices made in a **Coordination Game**, again with the simpler disc.



You picked the black sector. The other player picked the white sector. You therefore failed to coordinate and both of you earn £5 each.

The experiment consists of two periods. Each period consists of the three games as described above, using a 5-sector disc; the order of the games is random. At the end of the experiment one of the two periods will be randomly chosen. The earnings made in this period will be paid out in cash.

Again, please do not talk during this experiment! If you have questions just raise your hand.

Before the experiment there will be a quiz to check your understanding. Read hints carefully if you get stuck during the quiz.

### 3.E.2 Quiz

In this appendix you can find screenshots of the quiz which was conducted before the experiment. Participants who made a mistake in some part of the quiz were given a small hint and then were asked to repeat this part of the quiz.

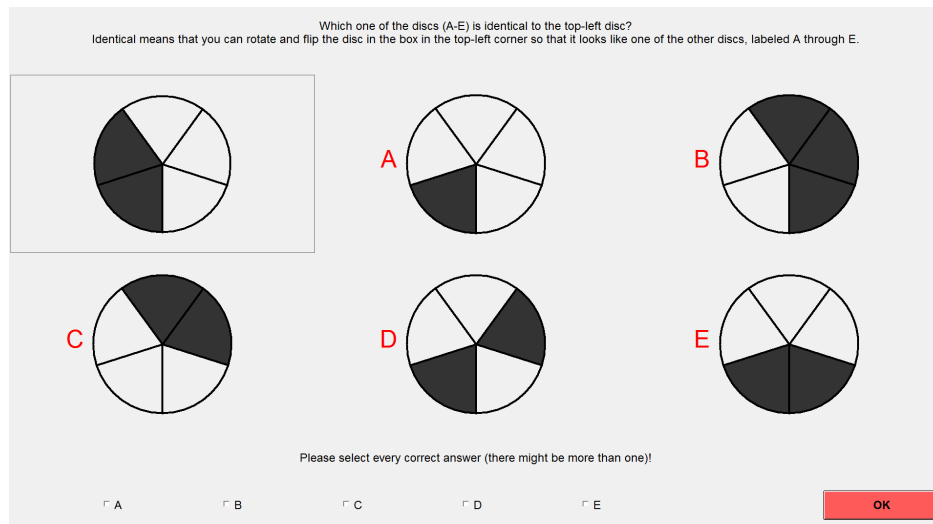


Figure 3.12: Quiz part 1

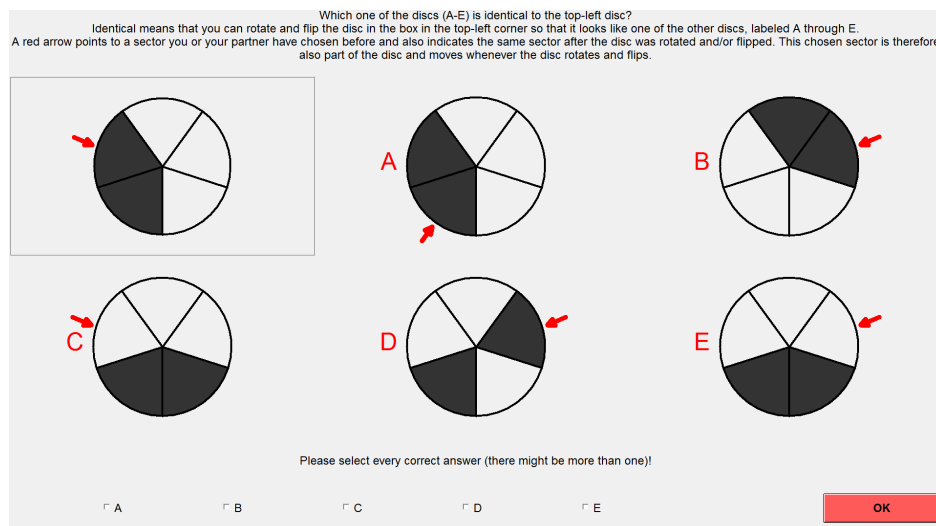


Figure 3.13: Quiz part 2

Assume you are playing the **Self Game** and you have chosen the same sector twice.

1) How much Pound will you get for this game?

Assume you and a random partner are playing the **Coordination Game** together and both of you have each picked a **different sector** of the same color.

2) How much Pound will you get for this game?

3) How much Pound will he get for this game?

Assume you are playing the **Guessing Game**. The participant you are observing is therefore playing the **Self Game**. He picked the same sector twice but you failed to pick the sector he picked.

4) How much Pound will you get for this game?

5) How much Pound will he get for this game?

Now imagine the same situation but this time you managed to pick the same sector he picked twice.

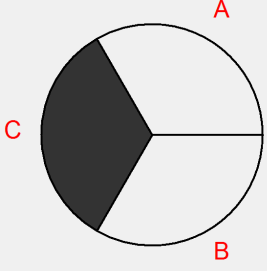
6) How much Pound will you get for this game?

7) How much Pound will he get for this game?

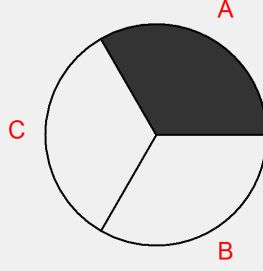
Figure 3.14: Quiz part 3

We are now going to test your knowledge of the **Self Game** as described in the instructions with an easier version of the disc. Remember, the labels (A-C) are not part of the disc.

**First Round :**



**Second Round :**



Assume you have chosen **sector C** (the black sector) in the first round (in the picture on the left). Where could your chosen sector be in the second round (in the picture on the right)? Please select every correct answer (there might be more than one)!

☐ A ☐ B ☐ C

Figure 3.15: Quiz part 4



We are now going to test your knowledge of the **Self Game** as described in the instructions with an easier version of the disc.  
Remember, the labels (A-C) are not part of the disc.

**First Round :**

**Second Round :**

Assume you have chosen **sector C** (one of the white sectors) in the first round (in the picture on the left).  
Where could your chosen sector be in the second round (in the picture on the right)?  
Please select every correct answer (there might be more than one!)

☐ A
 ☐ B
 ☐ C

OK

Figure 3.16: Quiz part 5

We are now going to test your knowledge of the **Guessing Game** as described in the instructions with an easier version of the disc.  
Assume you have observed the choices of the other player (which will not be possible in the experiment) and he has chosen a black sector twice).

Now you are presented this disc and you are playing the **Guessing Game**.

Assume you have chosen sector A (a white sector). What are your possible payoffs for this round?

☐ 0 Pounds  
☐ 2.5 Pounds  
☐ 5 Pounds

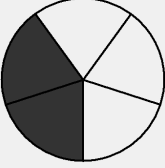
Now assume you have chosen sector B (the black sector). What are your possible payoffs for this round?!

☐ 0 Pounds  
☐ 2.5 Pounds  
☐ 5 Pounds

OK

Figure 3.17: Quiz part 6

We are now going to test your knowledge of the **Coordination Game** as described in the instructions. Assume you are playing this game with the five sector disc below.



Assume you know that your partner has chosen a **black sector**.

What is the probability of winning if you choose a black sector?

- ☐ 0% (no chance)
- ☐ 20% (a 1 in 5 chance)
- ☐ 33% (a 1 in 3 chance)
- ☐ 50% (a 1 in 2 chance)
- ☐ 100%

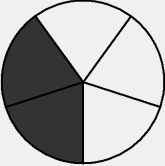
What is the probability of winning if you choose a white sector?

- ☐ 0% (no chance)
- ☐ 20% (a 1 in 5 chance)
- ☐ 33% (a 1 in 3 chance)
- ☐ 50% (a 1 in 2 chance)
- ☐ 100%

OK

Figure 3.18: Quiz part 7

We are now going to test your knowledge of the **Coordination Game** as described in the instructions. Assume you are playing this game with the five sector disc below.



Assume you know that your partner has chosen a **white sector**.

What is the probability of winning if you choose a black sector?

- ☐ 0% (no chance)
- ☐ 20% (a 1 in 5 chance)
- ☐ 33% (a 1 in 3 chance)
- ☐ 50% (a 1 in 2 chance)
- ☐ 100%

What is the probability of winning if you choose a white sector?

- ☐ 0% (no chance)
- ☐ 20% (a 1 in 5 chance)
- ☐ 33% (a 1 in 3 chance)
- ☐ 50% (a 1 in 2 chance)
- ☐ 100%

OK

Figure 3.19: Quiz part 8

## Chapter 4

## Bibliography

Marina Agranov, Elizabeth Potamites, Andrew Schotter, and Chloe Tergiman. Beliefs and endogenous cognitive levels: An experimental study. *Games and Economic Behavior*, 75(2):449–463, 2012.

Larbi Alaoui and Antonio Penta. Endogenous depth of reasoning. *The Review of Economic Studies*, page rdv052, 2015.

Carlos Alós-Ferrer and Christoph Kuzmics. Hidden symmetries and focal points. *Journal of Economic Theory*, 148(1):226–258, 2013.

Ayala Arad and Ariel Rubinstein. The 11–20 money request game: a level-k reasoning study. *The American Economic Review*, 102(7):3561–3573, 2012.

Robert J. Aumann and Michael B. Maschler. *Repeated Games With Incomplete Information*. MIT Press, Cambridge, MA, 1995.

Michael Bacharach. *Variable Universe Games*. MIT Press, 1993.

Sandeep Baliga and Tomas Sjöström. Arms races and negotiations. *The Review of Economic Studies*, 71(2):351–369, 2004.

Dirk Bergemann and Stephen Morris. Information design and bayesian persuasion. *American Economic Review*, Paper and Proceedings(106(5)):586–591, 2016a.

- Dirk Bergemann and Stephen Morris. Bayes correlated equilibrium and the comparison of information structures in games. *Theoretical Economics*, 11:487–522, 2016b.
- V Bhaskar. Egalitarianism and efficiency in repeated symmetric games. *Games and Economic Behavior*, 32(2):247–262, 2000.
- Cristina Bicchieri. *The grammar of society: The nature and dynamics of social norms*. Cambridge University Press, 2005.
- Andreas Blume. Coordination and learning with a partial language. *Journal of Economic Theory*, 95(1):1–36, 2000.
- Andreas Blume and Uri Gneezy. An experimental investigation of optimal learning in coordination games. *Journal of Economic Theory*, 90(1):161–172, 2000.
- Andreas Blume and Uri Gneezy. Cognitive forward induction and coordination without common knowledge: An experimental study. *Games and Economic Behavior*, 68(2):488–511, 2010.
- Adam Brandenburger and Eddie Dekel. Hierarchies of beliefs and common knowledge. *Journal of Economic Theory*, 59(1):189–198, 1993.
- Antonio Cabrales, Rosemarie Nagel, and Roc Armenter. Equilibrium selection through incomplete information in coordination games: an experimental study. *Experimental Economics*, 10(3):221–234, 2007.
- Colin F Camerer, Teck-Hua Ho, and Juin-Kuan Chong. A cognitive hierarchy model of games. *The Quarterly Journal of Economics*, pages 861–898, 2004.
- Hans Carlsson and Eric Van Damme. Global games and equilibrium selection. *Econometrica: Journal of the Econometric Society*, pages 989–1018, 1993.
- André Casajus. Focal points in framed strategic forms. *Games and Economic Behavior*, 32(2):263–291, 2000.

- André Casajus. Static focal points in framed strategic games. In *Focal Points in Framed Games*, pages 9–54. Springer, 2001.
- Michael Suk-Young Chwe. *Rational ritual: Culture, coordination, and common knowledge*. Princeton University Press, 2013.
- Russell W Cooper, Douglas V DeJong, Robert Forsythe, and Thomas W Ross. Selection criteria in coordination games: Some experimental results. *The American Economic Review*, 80(1):218–233, 1990.
- Camille Cornand. Speculative attacks and informational structure: an experimental study. *Review of International Economics*, 14(5):797–817, 2006.
- Giancarlo Corsetti, Amil Dasgupta, Stephen Morris, and Hyun Song Shin. Does one soros make a difference? a theory of currency crises with large and small traders. *Review of economic Studies*, pages 87–113, 2004.
- Miguel Costa-Gomes, Vincent P Crawford, and Bruno Broseta. Cognition and behavior in normal-form games: An experimental study. *Econometrica*, pages 1193–1235, 2001.
- Vincent P Crawford and Hans Haller. Learning how to cooperate: Optimal play in repeated coordination games. *Econometrica: Journal of the Econometric Society*, pages 571–595, 1990.
- Vincent P Crawford, Uri Gneezy, and Yuval Rottenstreich. The power of focal points is limited: Even minute payoff asymmetry may yield large coordination failures. *The American Economic Review*, pages 1443–1458, 2008.
- Vincent P Crawford, Miguel A Costa-Gomes, and Nagore Iriberri. Structural models of nonequilibrium strategic thinking: Theory, evidence, and applications. *Journal of Economic Literature*, 51(1):5–62, 2013.

- John Duffy and Jack Ochs. Equilibrium selection in static and dynamic entry games. *Games and Economic Behavior*, 76(1):97–116, 2012.
- Jeffrey C Ely. Beeps. *The American Economic Review*, 107(1):31–53, 2017.
- Urs Fischbacher. z-tree: Zurich toolbox for ready-made economic experiments. *Experimental economics*, 10(2):171–178, 2007.
- Sidney D Gamble. The disappearance of foot-binding in tingshien. *American Journal of Sociology*, pages 181–183, 1943.
- Matthew Gentzkow and Emir Kamenica. Costly persuasion. *The American Economic Review*, 104(5):457–462, 2014.
- Matthew Gentzkow and Emir Kamenica. Competition in persuasion. *The Review of Economic Studies*, 84(1):300–322, 2016a.
- Matthew Gentzkow and Emir Kamenica. A rothschild-stiglitz approach to bayesian persuasion. *The American Economic Review*, 106(5):597–601, 2016b.
- Matthew Gentzkow and Emir Kamenica. Bayesian persuasion with multiple senders and rich signal spaces. *Games and Economic Behavior*, 104:411–429, 2017a.
- Matthew Gentzkow and Emir Kamenica. Disclosure of endogenous information. *Economic Theory Bulletin*, 5(1):47–56, 2017b.
- David Gill and Victoria L Prowse. Cognitive ability, character skills, and learning to play equilibrium: A level-k analysis. *Journal of Political Economy*, 2016.
- Itay Goldstein and Ady Pauzner. Demand–deposit contracts and the probability of bank runs. *the Journal of Finance*, 60(3):1293–1327, 2005.
- John C Harsanyi. Games with incomplete information played by bayesian players part i-iii. *Management Science*, 14(5):320–334, 1967/68.

- John C Harsanyi and Reinhard Selten. *A general theory of equilibrium selection in games*, volume 1. The MIT Press, 1988.
- Frank Heinemann, Rosemarie Nagel, and Peter Ockenfels. The theory of global games on test: Experimental analysis of coordination games with public and private information. *Econometrica*, 72(5): 1583–1599, 2004.
- Christian Hellwig. Imperfect common knowledge of preferences in global coordination games. *UCLA Economics Online Papers*, 211, 2002.
- Teck-Hua Ho, Colin Camerer, and Keith Weigelt. Iterated dominance and iterated best response in experimental "p-beauty contests". *American Economic Review*, pages 947–969, 1998.
- Johannes Hörner and Andrzej Skrzypacz. Selling information. *Journal of Political Economy*, 124(6):1515–1562, 2016.
- Andrea Isoni, Anders Poulsen, Robert Sugden, and Kei Tsutsui. Focal points in tacit bargaining problems: Experimental evidence. *European Economic Review*, 59:167–188, 2013.
- Emir Kamenica and Matthew Gentzkow. Bayesian persuasion. *American Economic Review*, 101:2590–2615, October 2011.
- Margaret E Keck and Kathryn Sikkink. *Activists beyond borders: Advocacy networks in international politics*, volume 6. Cambridge Univ Press, 1998.
- Terri Kneeland. Identifying higher-order rationality. *Econometrica*, 83(5):2065–2079, 2015.
- Terri Kneeland. Coordination under limited depth of reasoning. *Games and Economic Behavior*, 2016.
- Anton Kolotilin, Tymofiy Mylovanov, Andriy Zapechelnyuk, and Ming Li. Persuasion of a privately informed receiver. *Econometrica*, forthcoming.

- Christoph Kuzmics, Thomas Palfrey, and Brian W Rogers. Symmetric play in repeated allocation games. *Journal of Economic Theory*, 154: 25–67, 2014.
- Robert Duncan Luce and Howard Raiffa. *Games and decisions: Introduction and critical surveys*. Wiley New York, NY., 1957.
- Gerry Mackie. Ending footbinding and infibulation: A convention account. *American sociological review*, pages 999–1017, 1996.
- Gerry Mackie and John LeJeune. Social dynamics of abandonment of harmful practices: a new look at the theory. *Special Series on Social Norms and Harmful Practices, Innocenti Working Paper*, 6, 2009.
- Judith Mehta, Chris Starmer, and Robert Sugden. The nature of salience: An experimental investigation of pure coordination games. *The American Economic Review*, 84(3):658–673, 1994.
- Jean-Francois Mertens and Shmuel Zamir. Formulation of bayesian analysis for games with incomplete information. *International Journal of Game Theory*, 14(1):1–29, 1985.
- Dov Monderer and Dov Samet. Approximating common knowledge with common beliefs. *Games and Economic Behavior*, 1(2):170–190, 1989.
- Stephen Morris and Hyun Song Shin. Unique equilibrium in a model of self-fulfilling currency attacks. *American Economic Review*, pages 587–597, 1998.
- Stephen Morris and Hyun Song Shin. Coordination risk and the price of debt. *European Economic Review*, 48(1):133–153, 2004.
- Rosemarie Nagel. Unraveling in guessing games: An experimental study. *The American Economic Review*, pages 1313–1326, 1995.
- John Nash. Non-cooperative games. *Annals of mathematics*, pages 286–295, 1951.



- Martin T Orne. On the social psychology of the psychological experiment: With particular reference to demand characteristics and their implications. *American psychologist*, 17(11):776, 1962.
- Eugenio Proto, Aldo Rustichini, and Andis Sofianos. Higher intelligence groups have higher cooperation rates in the repeated prisoner's dilemma. 2014.
- Ariel Rubinstein. The electronic mail game: Strategic behavior under "almost common knowledge". *The American Economic Review*, pages 385–391, 1989.
- Thomas C Schelling. *The strategy of conflict*. Harvard University Press, 1960.
- Dale O Stahl and Paul W Wilson. On players' models of other players: Theory and experimental evidence. *Games and Economic Behavior*, 10(1):218–254, 1995.
- Tomasz Strzalecki. Depth of reasoning and higher order beliefs. *Journal of Economic Behavior & Organization*, 108:108–122, 2014.
- Robert Sugden. A theory of focal points. *The Economic Journal*, pages 533–550, 1995.
- Wataru Tamura. A theory of multidimensional information disclosure, June 2014. URL <http://www.iser.osaka-u.ac.jp/library/dp/2012/DP0828.pdf>.
- Tostan. Female genital cutting. <http://www.tostan.org/female-genital-cutting>, 2016. Accessed: 2016-05-21.
- UNICEF. Female genital mutilation/cutting: a global concern. [http://www.unicef.org/media/files/FGMC\\_2016\\_brochure\\_final\\_UNICEF\\_SPREAD.pdf](http://www.unicef.org/media/files/FGMC_2016_brochure_final_UNICEF_SPREAD.pdf), 2016.
- John B Van Huyck, Raymond C Battalio, and Richard O Beil. Tacit coordination games, strategic uncertainty, and coordination failure. *The American Economic Review*, pages 234–248, 1990.

Jonathan Weinstein and Muhamet Yildiz. Impact of higher-order uncertainty. *Games and Economic Behavior*, 60(1):200–212, 2007a.

Jonathan Weinstein and Muhamet Yildiz. A structure theorem for rationalizability with application to robust predictions of refinements. *Econometrica*, 75(2):365–400, 2007b.

World Bank Group. *World Development Report 2012: Gender Equality and Development*. World Bank Publications, 2012.